CORRELATION BETWEEN PHYSICAL AND GENETIC MAPS OF PLASMIDS

IGOR Z. ZUBRZYCKI

Thesis Presented for the Degree of
DOCTOR OF PHILOSOPHY
in the Department of Biochemistry
UNIVERSITY OF CAPE TOWN
August 1992

ACKNOWLEDGMENTS

This work is a result of three years of study under the supervision of Professor Horst Klump to whom I would like to express my gratitude for his enthusaism, patient guidance, critical discussion and financial support.

I would like to thank Professor C.von Holt who made it possible to come to work at the University of Cape Town.

Special word of appreciation are owed to Dr Paul van Helden for his generous help during the completion of this work.

I am truly grateful to Dr H. Patterton for the generous gifts of the plasmids pGV403, pHP2 and pHP32 clones and complete sequences.

I would like to dedicate my thesis to my parents for giving me all their savings for air ticket to South Africa and for their great support during these three years.

I want also thank the Foundation for Research Development (FRD) for its financial assistance.

ABSTRACT

The aim of this study is to establish a correlation between the physical maps of plasmid DNA (in the form of calorimetric profiles, thermal denaturation profiles and electron micrographs of partly melted DNA sequences) and genetic maps of these DNAs and thus deal with questions which were not answered by previous researchers, viz:

Is there a correlation between base sequence function and a measurable physical property which can be assigned to biologically important units such as promoters or coding sequences?

Is there a correlation between the denaturation of gene sequences and cooperative transitions observed in a given temperature interval?

To answer these questions, a systematic study was initiated based on two families of plasmids with three different genes incorporated, namely the pGV403 family which contains a Chloramphenicol resistance gene on one side and the pUC9 family which contains an Ampicillin and a Tetracycline resistance gene on the other side.

Three different techniques were used to address the stated problems i.e. differential scanning calorimetry, high resolution thermal denaturation

and electron microscopy. The reason for using three techniques instead of only one or two as in previous studies is that each technique gives specific results which can be supplemented by the other techniques and only in this way it will be possible to approach a deeper understanding of changes induced by perturbing the sequence based structural integrity by elevating temperature.

In addition to measuring the experimentally observable parameters listed, a theoretical model was developed to predict the changes. This approach is termed local compositional complexity (LCC) analysis. The final goal of this investigation was to establish whether there is any correlation between the local compositional complexity and these selected genetic units.

Based on the calorimetric experiments an improved table of thermodynamic data including the stacking energy for ten different combinations of basepairs is presented. The prediction of a melting curve based on primary structure information can be based on the enthalpy individual combination of basepairs[41,50,58]. The tables of the thermodynamic data published in the literature are given in Appendix D. In this thesis a slightly different approach to predict t_m's was chosen (cf. p 67).

The results obtained by this combined approach showed that there is indeed a correlation between the specific base sequence of a given plasmid DNA and its biologically important units (genes) and thus confirms that there is a semiempirical correlation between genes and the observed cooperative melting units.

Table of Contents

LIST OF FIGURES

ABBREVIATIONS

ABBREVIATION	DESCRIPTION
bp	Base pair
DSC	Differential scanning calorimetry
EDTA	Ethylenediaminetetra-acetic acid
L broth	Luria broth
LCC	Local compositional complexity
EM	Electron Microscopy
0.1*SSC	0.015M NaCl+0.0015M trisodium citrate
Tris	Tris(hydroxymethyl)-aminomethane
ΔH	transition enthalpy
ΔH_{VH}	van't Hoff enthalpy
t_m	melting temperatures
F_{GC}	fraction of GC in a melting region
EtBr	Ethidium Bromide
TE-buffer	Tris-EDTA buffer
TBE-buffer	Tris-borate/EDTA buffer
BAC	benzyl dimethyl-alkyl-ammonium chloride

CHAPTER I

Introduction

Soon after the discovery of the DNA structure [3], the properties of the double stranded DNA helix were extensively studied by biologists, chemists and physicists. Once it was established that DNA was a complementary double helix, it was realized that the base sequence was the repository of genetic information and that the whole genome can be described as a kind of a mosaic structure [56]. This sparked the interest in the physical properties of the DNA molecule, represented by local base composition, and its biological function, represented by genes.

The interest in the physical properties of DNA sequences has focused on the features of the double helical folding (secondary structure) and on the correlation between the chemical composition (sequence), which is defined as a linear array of nucleotides and the mosaic of coding and non-coding regions. The secondary structure evolved to store and protect genetic information in the double helix structure.
The most direct physical approach employed to study sequence properties is to measure physical parameters which change with temperature.

The helix to coil transition of DNA (a process during which strand separation occurs due to break down of all H-bonds) has been the subject of numerous reviews [11,14], and it has been known for many

years that many important biological processes such as replication, transcription or recombination involve disruption of the duplex structure of DNA. It is clear that understanding the mechanism of strand separation is of fundamental importance.

Strand separation may be caused by chemical, physical or enzymatic perturbation (e.g. pH, elevated temperature, RNA polymerase) and is a function of the base composition of the DNA. Experimentally this process can be monitored by spectroscopic methods such as high resolution thermal denaturation [52]. Recording changes in ***UV*** absorbance while increasing the temperature has become one of the most popular experimental techniques because of the straightforward correlation between the temperature change and the extent of melting of the DNA molecule and because of the accurate definition of the experimental conditions that is possible.

The ***UV*** absorbance of purine and pyrimidine bases is known as caused by an π-π^* electronic transition. The absorbance intensity increases about 30 to 40% during the release of the bases from their double helical stacking (degeneration).

A thermal denaturation profile (change in absorbance versus temperature) is characterized by its t_ms (the melting temperature i.e. the temperature at which 50% of the total absorbance change occurs). As a result of many studies of the melting profiles of various DNA sequences, a linear relationship between the melting temperature and the GC content has been noted (GC base-pairs melt at higher temperatures than AT base-pairs). Thus the melting temperature has become a valuable tool for measuring the GC content of DNA molecule [4]. This type of result

implies that a region within a DNA sequence with higher percentage of AT base-pairs has a lower thermal stability than a GC rich area. It is interesting to note that promoters (of genes) are frequently located in or adjacent to regions of DNA that contain a high percentage of AT base pairs [24,26,38]. This richness in AT may facilitate the opening of the helical structure (melting) and promote transcription [17,25].

It was discovered quite early that the melting of certain DNA sequences does not occur in a single transition but occurs in discrete steps [15,18,27]. Thermal denaturation studies in recent years have illustrated this convincingly and a series of fine structure stability maps of various DNA sequences [12,13,15,19,28,37] has been published.

In 1976 Gotoh *et. al.* [19] published a most informative paper concerning the thermal stability of bacteriophage lambda DNA where the melting profile of λc I857S7. DNA was compared with λc Ib2b5 DNA. The result of this comparison shows the disappearance of certain peaks in λc I857S7 as compared to λc Ib2b5 DNA (cooperative melting regions). This showed that λc Ib2b5 mutants with deleted DNA regions had lost some thermodynamically important cooperative units. (The word "peak" will be used throughout this thesis to describe differential maxima in thermal denaturation for ease of writing and reading).

In 1978 Tachibana *et. al.*[33] used high-resolution thermal denaturation to map the whole genome of fd phage. They examined the overlapping

fragments obtained by restriction enzyme digestion of the full length DNA and the complete linear form of the phage DNA. The objective of this study was to map regions of the genome according to regional thermal stability and to compare these regions with the genetic map of the same genome. The results suggested that indeed some correlation exists between genes and cooperative melting regions. This was confirmed by other groups[35,38].

One of the most recent studies using this type of approach was published in 1979 [37]. The authors examined which restriction enzyme digest fragments of known sequence composition carry genetic information from the lactose operon. The result of their work represents an advance in so far as it shows an attempt to compare a defined sequence with the relevant melting profile (peak).

The results of the previous studies allow one to classify melting profiles into three basic categories according to the size and homogeneity of the genome: (1) The polyphasic melting of short homogeneous viral DNA (2) the smooth and monotonic melting of bacterial DNAs, and (3) the broad melting profile for DNAs of higher eukaryotes. Based on the previous results [33,19,20] a cooperative melting unit was defined. The size of the melting unit which appears as a single peak (transition) was established to be 300 to 1500 base pairs for the ΦX174 DNA, 100 to 1100 base pairs for the fd phage and 200~500 to 1500~3000 of λ phage. This finding made it possible to define a theoretical formula for the calculation of the length of the melting unit: The length of the melting unit corresponding to a single transition can be estimated to be a product of the length of

the entire DNA molecule multiplied by a factor. The factor is calculated from the ratio of the area under the peak (transition) to the total area under the melting profile.

Felsenfeld and Hirschman [8] showed that melting starts from AT-rich regions and as the temperature rises propagates to more GC-rich regions (this correlates well with results obtained by Marmur and Doty [4]).

Differential scanning calorimetry (***DSC***) is an accurate alternative method to UV-melting for physically mapping melting regions along DNA molecules. It is a useful tool to establish thermodynamic parameters such as enthalpy, entropy and free energy of helix to coil transition of the DNA molecule [10,21,22,33,34,34,36]. Most of the calorimetric profiles reported in the literature were obtained from studies on (large) chromosomal DNA which revealed broad heat absorption profiles.

However, Y.Maeda *et. al.* [46] obtained multimodal transition profiles when they used the Chloramphenicol resistance plasmid **pJL3-TB5** DNA. *DSC* scans of this plasmid DNA showed six transitions (peaks) over the temperature range 82°C to 98°C. Since this plasmid has been characterized, it was possible to assign each particular peak to a given region with a defined GC content.
Derivatives (mutants) of this plasmid were also studied [41] and it was observed that the number of peaks in the ***DSC*** scan depended on the sequence of mutants (generally the mutants showed fewer peaks than

the parental plasmid). Similar results for high resolution thermal denaturation were reported earlier by Gotoh *et al.* [19].

One of the most advanced calorimetric studies on plasmid DNA was done by Maeda and Othsubo [53]. They showed that the helix-coil transition of particular regions or "genes" of the plasmid **ColE1** resulted either in a single peak or in more in than one peak, depending on the length of those "genes".
As mentioned earlier, differential scanning calorimetry can be very useful for the acquisition of thermodynamic parameters of the purine and pyrimidine bases. The (ΔH_i) transition enthalpy, (ΔS_i)-entropy, and (ΔG_i) -free energy are related by the equation $\Delta G_i=\Delta H_i-T\Delta S_i$. In recent years three different papers which deal with this problem have appeared [41,47,53] and three different sets of thermodynamic data for all possible combinations of base pairs were reported. In this thesis a fourth set based only on calorimetric data (Appendix D) is presented.

Unfortunately both methods, differential scanning calorimetry as well as high resolution (UV) thermal denaturation, involve sizeable amounts of experimental material (ca 250 µg) which is a problem with this research.

A third method which is suitable for physical mapping of genes and which requires less material is electron microscopy (EM).
An important improvement of this methodology was the discovery that glyoxal stabilizes partially denatured regions of DNA [21] by interaction with GC basepairs (guanosine nucleotides). This technique was used by

Borovic *et. al.* [39] in the visualization of denaturation of ColE1 DNA. The results obtained by EM confirm those obtained by differential scanning calorimetry and UV-spectroscopic methods. There is a correlation between the appearance of sharp peaks in the differential melting curve or calorimetric curve and the melting of one or several DNA regions within the plasmid sequence.

The most recent paper dealing with this subject was published by Kalambet *et. al.* [48]. The authors compared differential melting curves of two linearized plasmids, **pA03** and **pBR322**, under a variety of ionic strengths with melting maps obtained from EM. The experimental results confirm the correspondence of peaks in the melting profile and denatured sequences on the EM map [39,21]

In summary it can be stated that the physical properties of DNA double helices can be studied by a variety of techniques, such as high resolution thermal denaturation, differential scanning calorimetry, and electron microscopy. Among the three methods only EM (electron microscopy) allows one to "freeze" physical changes by chemical reaction, but all techniques give unified results: The physical structure of the genome can be treated as a mosaic of physically and/or genetically important regions.

In this sense the thesis is a continuation of the previous experimental work (see above). It is an attempt to answer the following questions which remained unanswered by previous authors, viz:

1. Is there a correlation between the base sequence and a measurable physical property which can be assigned to biologically important units such as promoters or coding sequences?

2. Is there a correlation between the denaturation of gene sequences and cooperative transitions observed in a given temperature interval?

To answer these two fundamental questions the following approaches are suggested:

(1) DNA sequences can be considered as a text of a"language" which consist of "words" [60], where the "word" is composed of nucleotides such as triplets of basepairs, or
(2) DNA can be viewed as a mosaic of coding and noncoding regions where the smallest element is the cooperative melting unit.

As mentioned above base triplets are the elementary "words" of the genetic "language" while single bases are the elements of mutation. One can ask the following question: Is there a detectable influence of the simple "word" on the physical properties of DNA molecule? The experiments referred to in the introduction suggest that the answer to this question is actually "no". Any observable unit with impact on the physical properties must be of a larger size.

It was shown previously that melting units have an average length of 100~3000 base pairs for different DNAs. This corresponds to the average exon length as shown by Gilbert [62]. Thus 100 bp can be a very convenient unit for the "mosaic " theory of the genome, where 100 bp units represent building blocks. Longer cooperative units are simply made up by multiplications of the smallest element.
To demonstrate this in detail I have introduced the mosaic model of the DNA molecule to answer the two questions raised.

To gather the necessary experimental data to address the problem in this thesis a combination of three experimental methods i.e. differential scanning calorimetry (***DSC***), high resolution thermal denaturation and electron microscopy (***EM***), was used, while in all previous studies published in the scientific literature only one or at most two techniques were applied. The goal of this study is to extend our knowledge into this area of DNA research by comparing the results obtained by three different methods which are used to approach the same questions from three different directions. High resolution thermal denaturation studies measure mainly the base-stacking interaction, differential scanning calorimetry measures total enthalpy changes due to helix to coil transition reflecting base pairing and base stacking, and finally electron microscopy can visualize "frozen physical stages" induced by elevated temperatures. None of the methods alone is sufficient to give a correct and complete answer.
Only a combination of these methods can give a full and clear picture of the physical changes occurring during the melting process caused by elevated temperature.

In addition careful selection of experimental material was necessary. Experiments were carried out on three families of plasmid DNAs, where each family has its well defined "core" gene. This approach of selected sequences and appropriate techniques seemed to be the best solution to the problems stated above because it systematizes the research.

The thesis is divided into nine chapters. Chapters II, III and VI describe more extensively the main experimental methods i.e differential scanning calorimetry, high resolution thermal denaturation and electron microscopy, and the results obtained for the sets of two families of plasmid DNAs from **pGV403** and from **pUC9**. As an experimental approach to the "mosaic" theory of the DNA sequence, chapter VII presents "Thermal denaturation studies of genes" where genes are viewed as "building blocks" of the DNA "mosaic".
A synopsis of results and discussion is presented in chapter VIII.
Chapters IV and V contain a semiempirical approach to the prediction of sequence stability and loop energy.

CHAPTER II

Differential scanning calorimetry of plasmid DNA

Introduction

The conformational change of DNA from a helix to a random coil can be monitored by following the absorbance change (260 nm) as function of the temperature as well as by recording the excess heat capacity (ΔC_p) with the help of differential scanning calorimetry (***DSC***).

The aim of scanning microcalorimetry [16] is to measure the heat capacity difference between a reference (standard liquid) and a sample (sample diluted into the same solvent) to cancel all contributions to ΔC_p due to the solvent. Determination of the heat capacity difference requires first the determination of the ΔC_p of the standard liquid for the whole temperature range of the instrument to establish the instrument base line. Once the base line is established the sample can be loaded and the heat capacity over the whole temperature interval is recorded. The deviation of the experimental scan from the base line corresponds to the excess heat capacity ΔC_p of the sample studied. To calculate the transition enthalpy ΔH for the helix coil transition between the temperature T_i and the final temperature T_f the following equation is used.

$$\Delta H = \int_{T_i}^{T_f} \Delta C_p dT$$

where ΔC_p is measured excess heat capacity at constant pressure. It is known that in some DNA sequences more that one transition (peak) may

occur. The gross transition can be deconvoluted into a series of subtransitions generally occurring at distinctly different temperatures. It has been shown [47,53] that plasmid DNAs in particular give characteristic ***DSC*** curves composed of a series of endothermic peaks due to helix-coil transitions characteristic of each DNA sequence.

The goal of this part of the thesis is to show the correspondence between ***DSC*** curves of plasmids with similar base sequences (fingerprinting) to demonstrate that systematic alterations between sequences effect specific subtransitions in the melting curves.

To this aim calorimetric scans were performed with DNA from two groups of plasmids
(1) pGV403, **pHP32** and **pHP2**, and **(2) pUC9**, **pBS** and **pBR322** respectively, which vary characteristically in their nucleotide sequence within each group.
Details of the sequence of each plasmid (parent plasmid + insert) are provided in the Appendices A and B.

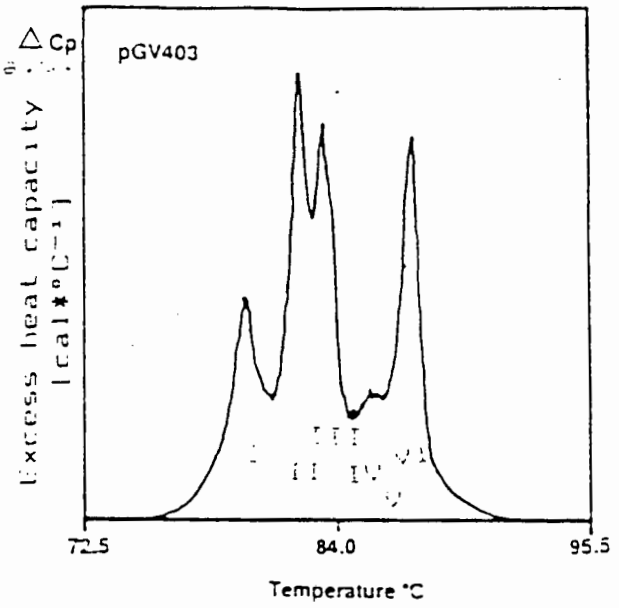
II.1. Recording and representation of DSC melting profiles of the linearized plasmids pGV403 pHP2 and pHP32

The ***DSC*** curves for the melting of **pGV403** and its derivatives were scanned at a heating rate of 1°C per minute in standard buffer (50mM NaCl, 0.5mM EDTA, 1mM Tris, pH 7.6) and are presented in Figures 1a-c. The melting curves of plasmid **pGV403** show six endothermic transitions over the temperature range from 79°C to 90°C. A similar set of transitions occurs in the melting profile of plasmid **pHP2** over the temperature range from 78°C to 90°C while plasmid **pHP32** shows seven transitions over the same temperature range.

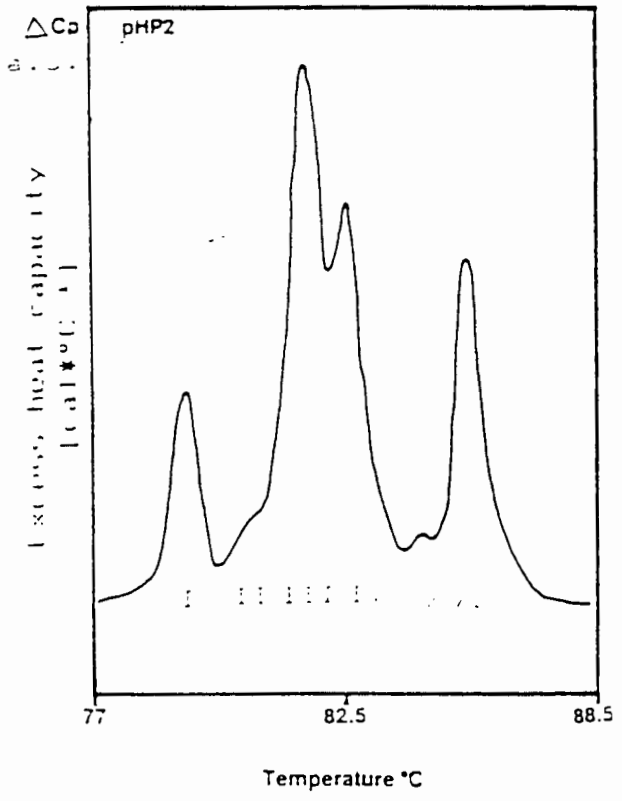
Figures 1d and 1e show the correspondence of the sequences of plasmid **pHP2** and **pHP32** respectively (based on the circular maps of these plasmids: Appendix A), and the sequence of plasmid **pGV403**.

It is obvious from the DSC profiles and sequences comparison profiles that differences between the denaturation profiles reflect differences between the sequence of these plasmids.

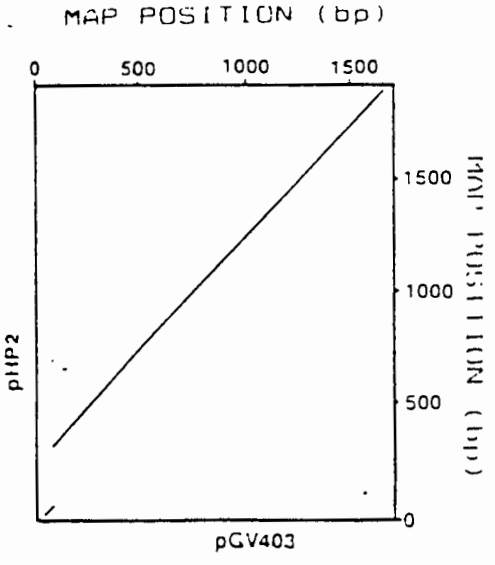
Fig 1a


ΔCp
pGV403
72.5
84.0
95.5
Temperature °C

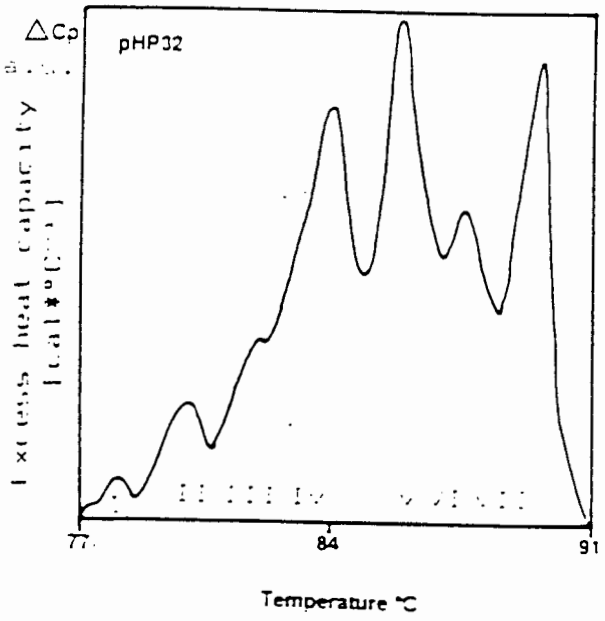
1b


ΔCp
pHP2
77
82.5
88.5
Temperature °C

1d


MAP POSITION (bp)
0
500
1000
1500
pHP2
pGV403

1c



1e

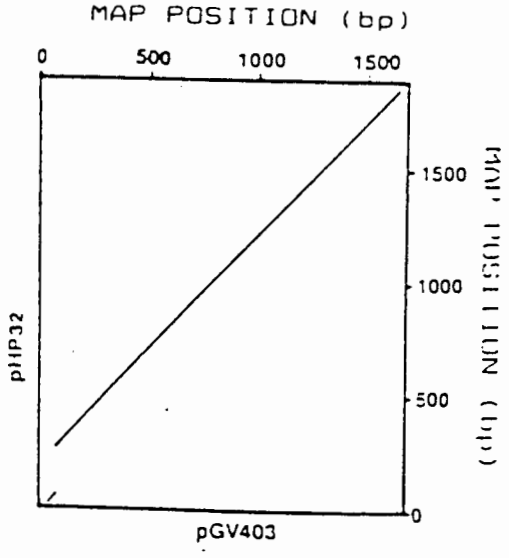


Fig 1. a) *DSC* melting profile of the relaxed circular DNA from plasmid **pGV403** b) *DSC* melting profile of the relaxed circular DNA from plasmid **pHP2** and comparison of sequences with **pGV403** (Fig 1d) c) *DSC* melting profile of the relaxed circular DNA from plasmid **pHP32** and comparison of sequences with **pGV403** (Fig 1e). DNA concentration used was 236µg per 0.5ml. Buffer: 50mM NaCl, 1mM Tris, 0.5mM EDTA pH 7.6. Sequence comparison was done as described elsewhere (Appendix B)

It can be seen that the basic melting profile of the **pGV403** DNA is conserved within the melting profiles of **pHP2** DNA and **pHP32** DNA like a fingerprint. Differences in the overall shape of the envelope curve were a true reflection of inserted sequences (*foreign* DNA) into the sequence of "parental" plasmid **pGV403**. From Figs 1 a-c it is obvious that each transition (peak) occurs at a well-defined temperature. The various t_ms of the characteristic peaks are summarized in Table 1.

Table 1

Number of peak	Temperature [°C]		
	pGV403	pHP2	pHP32
I	79.4	80.0	79.6
II	81.8	81.0	81.3
III	83.3	82.0	82.1
IV	85.4	83.2	83.4
V	86.5	85.5	85.9
VI	87.3	86.5	87.2
VII			89.4

Temperature reading error is +/- 0.1°C

II.1.a. Deconvolution of the DSC melting profiles of the linearized plasmid pGV403 and its derivatives

Melting profiles of DNAs from **pGV403**, **pHP32** and **pHP2** were deconvoluted using a Gaussian approximation of profiles (Fig 2a-c) [61]. Using a Gaussian curve to fit subtransition is sufficiently accurate for the evaluation of the thermodynamic data i.e. each transition enthalpy per subtransition and the corresponding transition temperature. However there are other approaches in use [35]. The description of the native/denatured equilibrium in nucleic acids requires the introduction of the partition function Q defined as a sum of the statistical weights of all possible states of the molecule

$$Q = \Sigma\, e^{-\Delta Gi/RT}$$

where ΔG_i is the Gibbs free energy of the i-th state, R the gas constant and T the absolute temperature.

Evaluation of this function requires the identification and enumeration of relevant states namely, native and denatured DNA conformations and their Gibbs free energies. The population of any particular state P_i is then given by the equation:

$$P_i = e^{-\Delta Gi/RT}/Q$$

and can be used to calculate observed system properties. Of particular importance is the average excess enthalpy function $<\Delta H>$ allowing for the calculation of the excess heat capacity function which can also be measured by differential scanning calorimetry:

$$<\Delta H> = \Sigma\, P_i \Delta H_i, \quad <C_p> = \delta<\Delta H>/\delta T$$

The area under each peak corresponds to the transition enthalpy of that particular transition. The quantitative result is derived by comparing the area to the electrically induced calibration mark (not shown) and listed in Table 2.

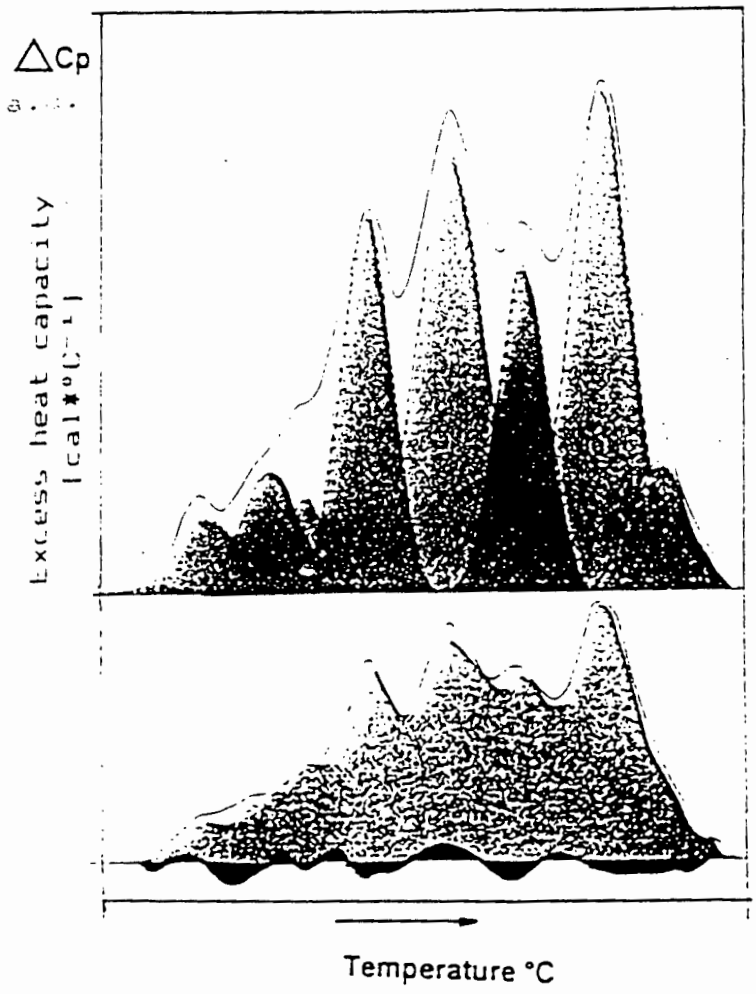
Table 2

Enthalpy (ΔH_{CAL}) of transitions obtained from DSC

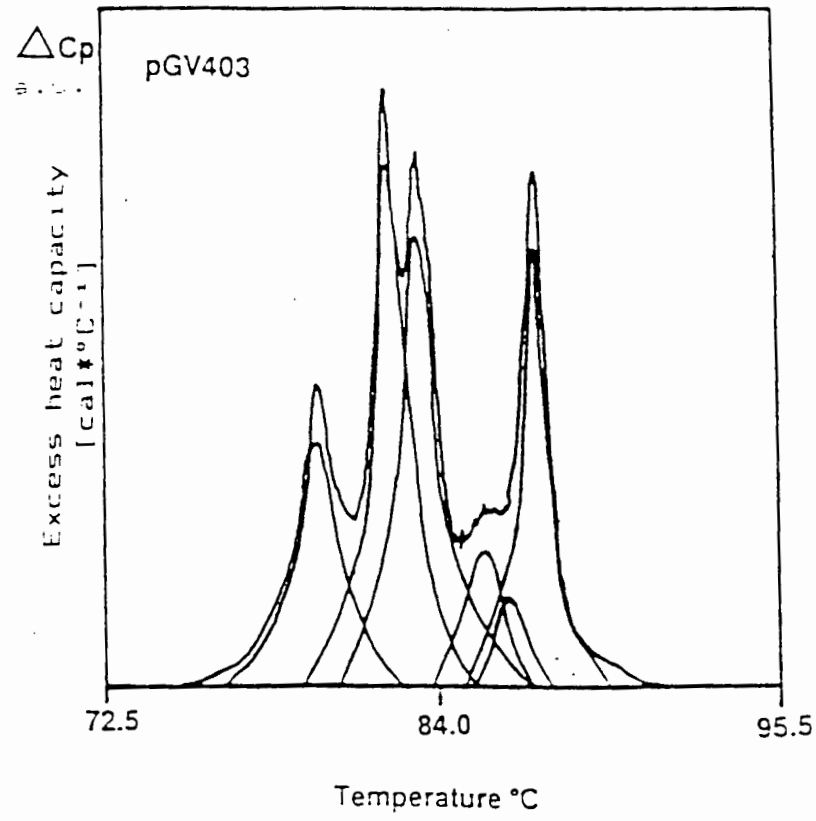
plasmid	Total ΔH [kJ/mol]*10^4	ΔH_{CAL} [kJ/mol]*10^3 Transition						
		1	2	3	4	5	6	7
pGV403	4.75	3.6	14.3	10.2	3.0	1.4	11.2	
pHP2	5.54	3.8	2.0	16.7	12.1	1.0	9.3	
pHP32	5.42	3.3	4.2	5.3	10.0	12.8	7.2	14.0

The scheme of the deconvolution of the differential scanning callorimetry curve is given in Fig 2. The upper part of the figure presents the deconvolution of the envelope curve (solid line) by a series of separate gaussian curves. Each bellshaped gaussian curve represents the melting of a cooperative unit, the peak maximum represents the melting temperature. In the lower part the superposition of the contribution of the cooperative units to the experimental transition curve is depicted. The deviation from the perfect match at any given temperature is shown as the small black peak on top of and below the zero line.

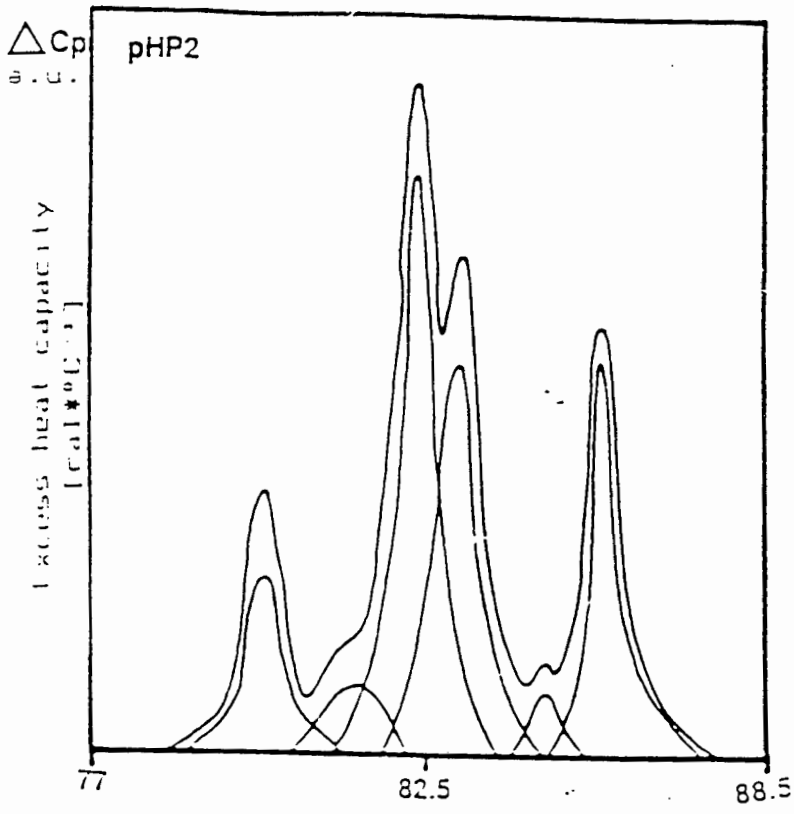
Fig 2


ΔCp
Excess heat capacity
Temperature °C

2a


ΔCp
pGV403
Excess heat capacity
72.5
84.0
95.5
Temperature °C

2b



2c

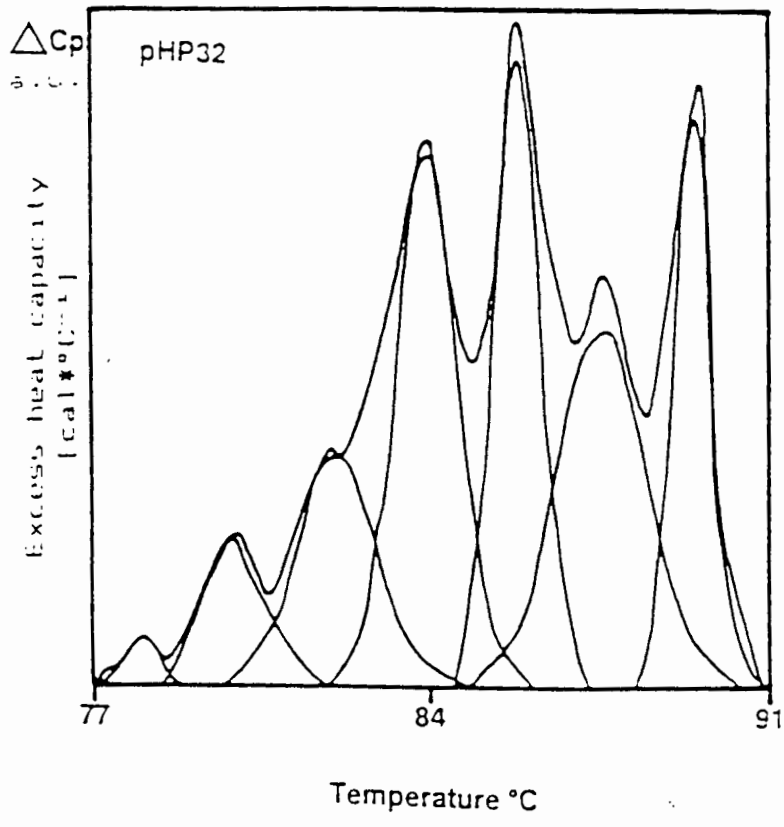
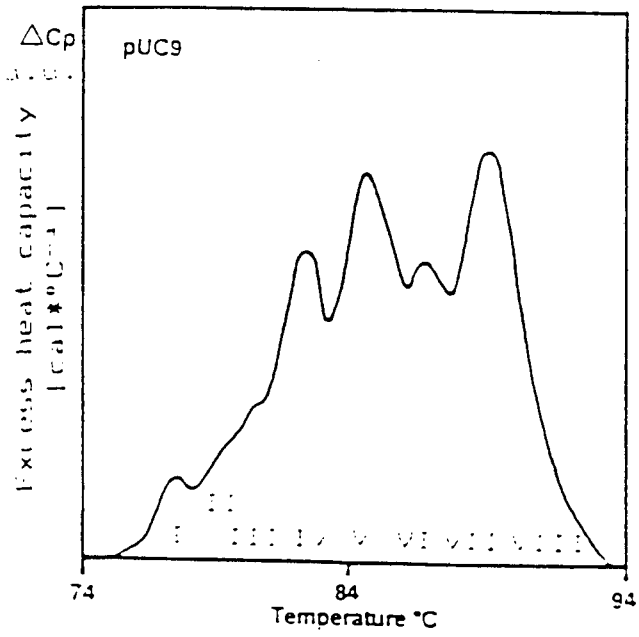


Fig 2. *DSC* curves deconvoluted using Gaussian approximation a) **pGV403** b) **pHP2** c) **pHP32**

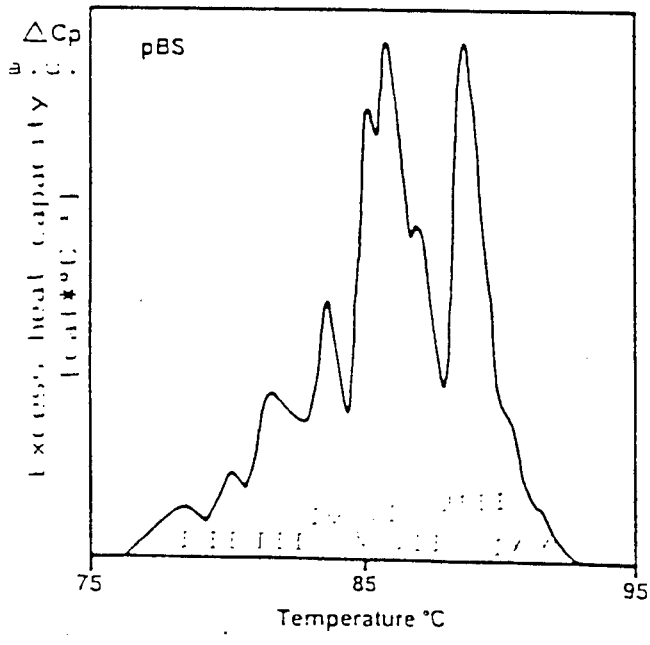
II.2. DSC melting profiles of the linearized plasmids pUC9, pBS and pBR322.

The melting profiles of **pUC9** DNA were obtained under the same experimental conditions as used for DNA from plasmid **pGV403**. Figs 3a-c show the melting profiles and the sequence comparison (Fig 3d,e) between **pUC9**, **pBS** and **pBR322** respectively. The circular sequences of these plasmids are presented in Appendix A

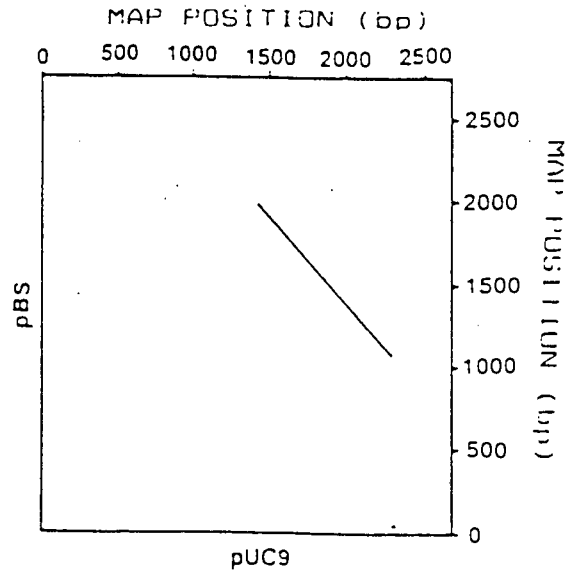
Fig 3a


ΔCp
pUC9
74
84
94
Temperature °C

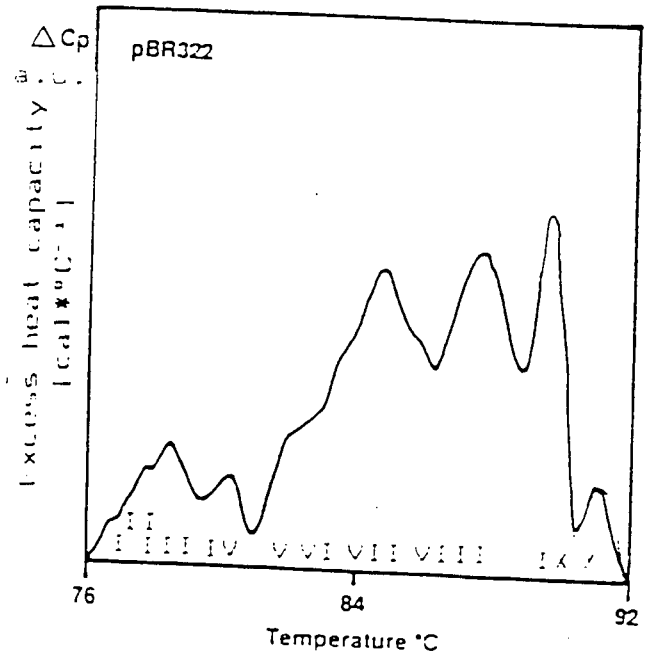
3b


ΔCp
pBS
75
85
95
Temperature °C

3d


MAP POSITION (bp)
0
500
1000
1500
2000
2500
pBS
pUC9

3c



3e

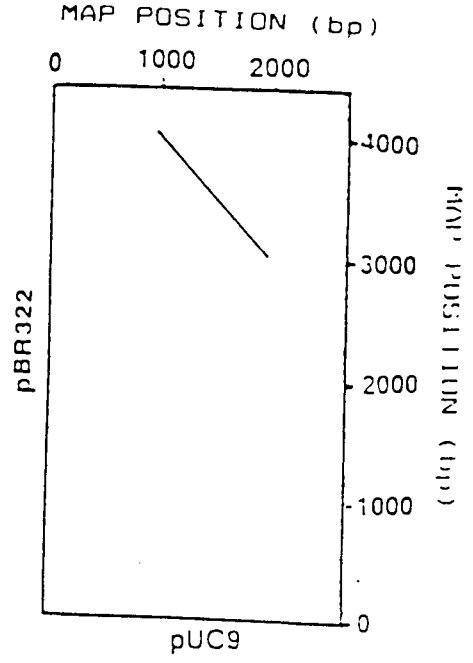


Fig 3. a) *DSC* melting profile of the relaxed circular DNA from plasmid **pUC9** b) *DSC* melting profile of the relaxed circular DNA from plasmid **pBS** and comparison of sequences with **pUC9** (Fig 3d) c) *DSC* melting profile of the relaxed circular DNA from plasmid **pBR322** and comparison of sequences with **pUC9** (Fig 3e). DNA concentration used was 236µg per 0.5ml. Buffer: 50mM NaCl, 1mM Tris, 0.5mM EDTA pH 7.6. Sequence comparison was done as described elsewhere (Appendix B)

The melting profile of **pUC9** DNA (3a) exhibits eight clearly separated transitions ("peaks") within the temperature range from 76°C to 90°C. The corresponding profile of plasmid **pBS** shows ten peaks within the temperature range from 77°C to 93°C and **pBR322** DNA gives the same number of peaks within the temperature range from 76°C to 91°C. The overall shape of this series of curves was not as conserved as that of the previous group of plasmids owing to a larger difference in base composition and the longer sequences. Particularly marked differences in shape can be observed between the profiles of the plasmid **pBR322** and **pUC9**. This may be due to the greater length of **pBR322** as compared to **pUC9** (1.6:1) and to considerable differences in nucleotide sequences. The t_m's of the individual peaks are listed in Table 3.

Table 3

DSC transitions of plasmids

t_ms of deconvoluted peaks from the melting profiles of plasmid **pUC9**, **pBS** and **pBR322**

Number of transition	Temperature[°C]		
	pUC9	pBS	pBR322
I	76.9	77.0	76.5
II	78.3	78.9	78.1
III	81.4	81.8	78.9
IV	82.9	83.3	80.8
V	84.4	85.4	82.7
VI	86.3	87.3	83.5
VII	88.5	89.0	84.5
VIII	89.2	89.8	85.6
IX		92.0	87.8
X		92.7	90.2

Temperature reading error is +/- 0.1°C

II.2.a. Deconvolution of the DSC melting profiles of the linearized plasmid pUC9 and its derivatives.

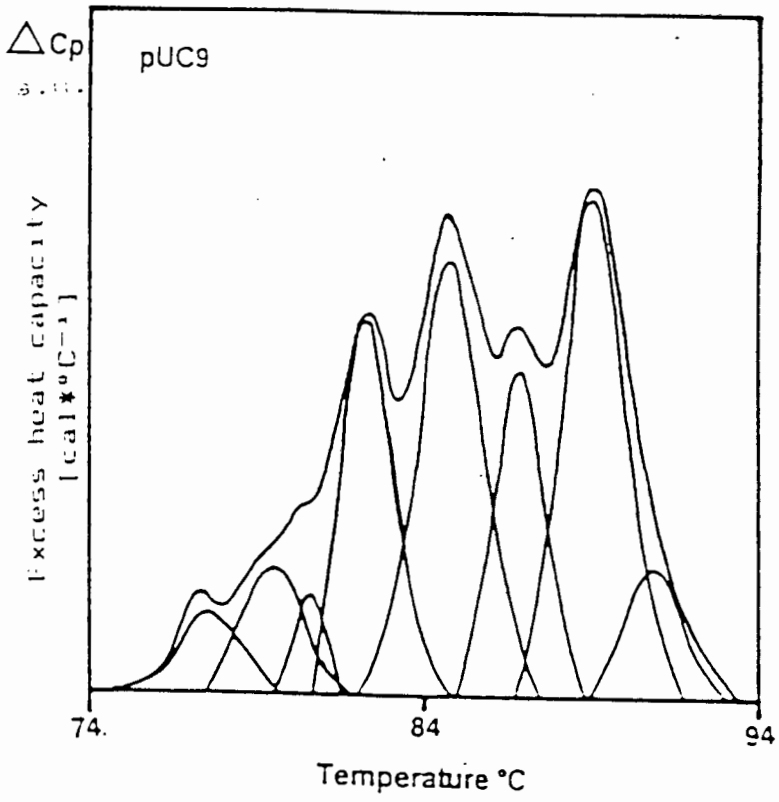
Melting profiles of DNAs from **pUC9**, **pBS** and **pBR322** were deconvoluted using a Gaussian approximation of profiles (Fig 4a-c). The area under each peak corresponds to the transition enthalpy of that particular transition. The quantitative result is derived by comparing the area to the electrically induced calibration mark (not shown) and listed in Table 4.

Table 4

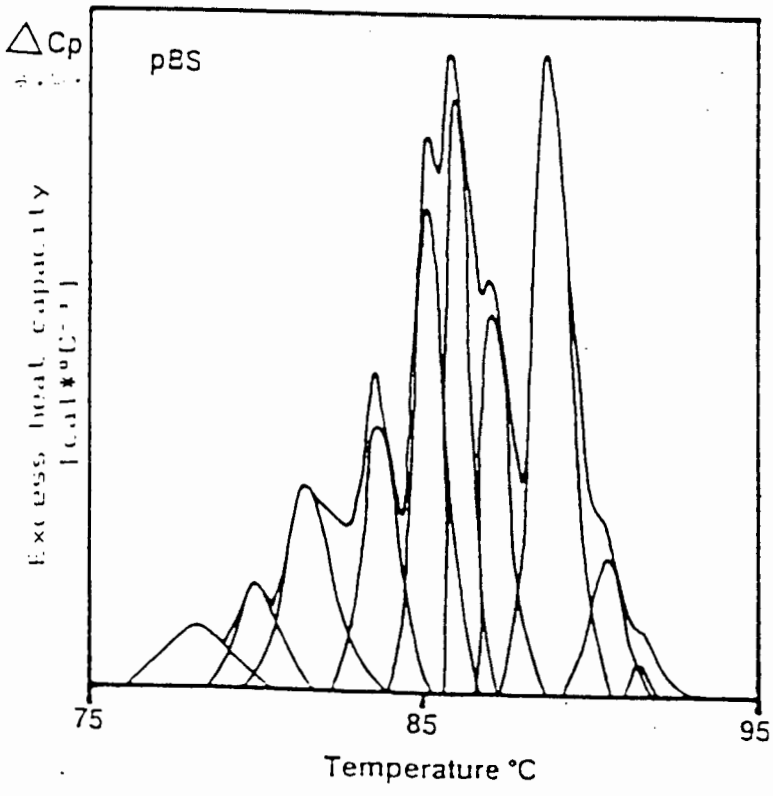
ΔH transition from DSC

no of peak	ΔH_{CAL} [kJ/mol]		
	pUC9	pBS	pBR322
ΔH total	$9.2*10^4$	$9.6*10^4$	$15.1*10^4$
I	$4.2*10^3$	$3.8*10^3$	$2.7*10^3$
II	$5.3*10^3$	$4.6*10^3$	$5.0*10^3$
III	$4.0*10^3$	$8.6*10^3$	$6.6*10^3$
IV	$1.4*10^4$	$1.0*10^4$	$4.9*10^3$
V	$2.2*10^4$	$1.3*10^4$	$1.0*10^4$
VI	$1.2*10^4$	$1.4*10^4$	$8.6*10^3$
VII	$2.4*10^4$	$1.3*10^4$	$3.2*10^4$
VIII	$5.4*10^3$	$2.5*10^4$	$4.8*10^4$
IX		$4.8*10^3$	$2.2*10^4$
X		$7.9*10^2$	$3.3*10^3$

Fig 4a


ΔCp
pUC9
Excess heat capacity
74.
84
94
Temperature °C

4b


ΔCp
pBS
75
85
95
Temperature °C

4c

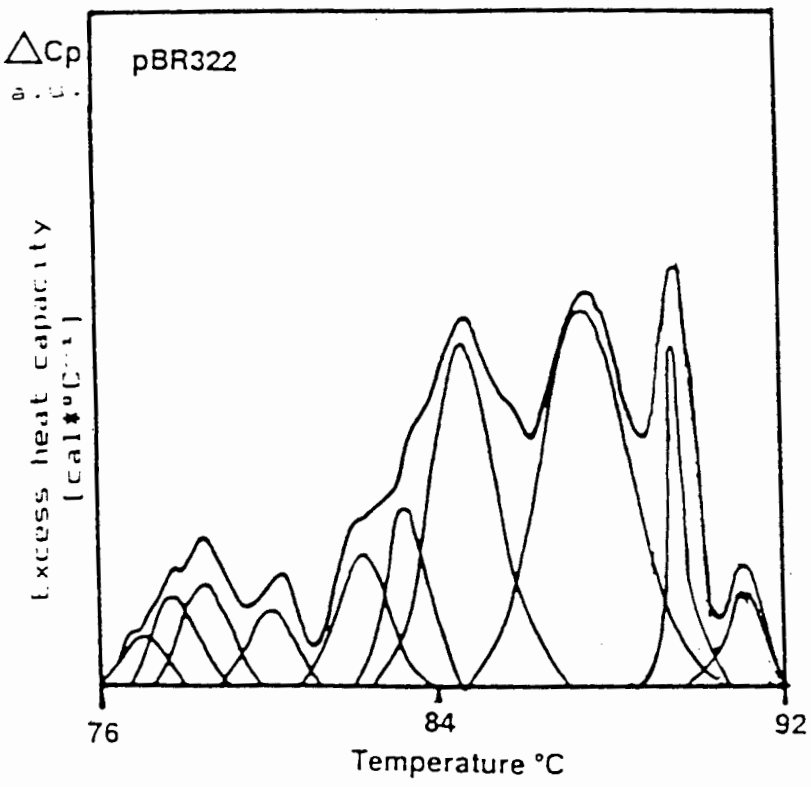


Fig 4. *DSC* profiles deconvoluted by Gaussian approximation
a) **pUC9** b) **pBS** c) **pBR322**

Summary and conclusion of the DSC experiments

The results obtained by differential scanning calorimetry presented here show convincingly that both groups of plasmids reveal similar melting profiles within their groups. It is obvious that the parental plasmid DNAs are reflected in the melting profile of the derived sequences, which exhibit extra features due to the sequence variations.

So it does not come as a surprise that the transitions occur at almost exactly the same temperature and have very similar ΔH values across both group of plasmids (cf. Table 5 and 6). To illustrate this further peak I of **pGV403** is compared to peak I of **pHP2** and peak I of **pHP32** respectively. The corresponding data are as follows.

Peak number I of **pGV403** (t_m 79.4°C,ΔH=3.6*10^3 kJ/mol) corresponds to peak number I of **pHP2** (t_m 80.0°C, ΔH=3.8*10^3 kJ/mol) and peak number I of **pHP32** (t_m 79.6°C, ΔH=3.3*10^3 kJ/mol).
A similar comparison can be made for peak III of **pGV403** and the corresponding peaks in the profiles of **pHP2** and **pHP32** (vide infra).

Peak number III of **pGV403** (t_m 83.3°C, ΔH=10.2*10^3 kJ/mol) corresponds to peak number IV of **pHP2** (t_m 83.2°C, ΔH=12.1*10^3 kJ/mol) and peak number IV of **pHP32** (t_m 83.4°C, ΔH=10.0 *10^3 kJ/mol).

From these examples one is tempted to conclude that similar sequences which exist in a variety of plasmids are detectable by corresponding thermal transition.

As a support to this claim a linear comparison of the entire sequences of **pGV403**, **pHP2** and **pHP32** is presented based on maps presented in Appendix B.

Fig 5

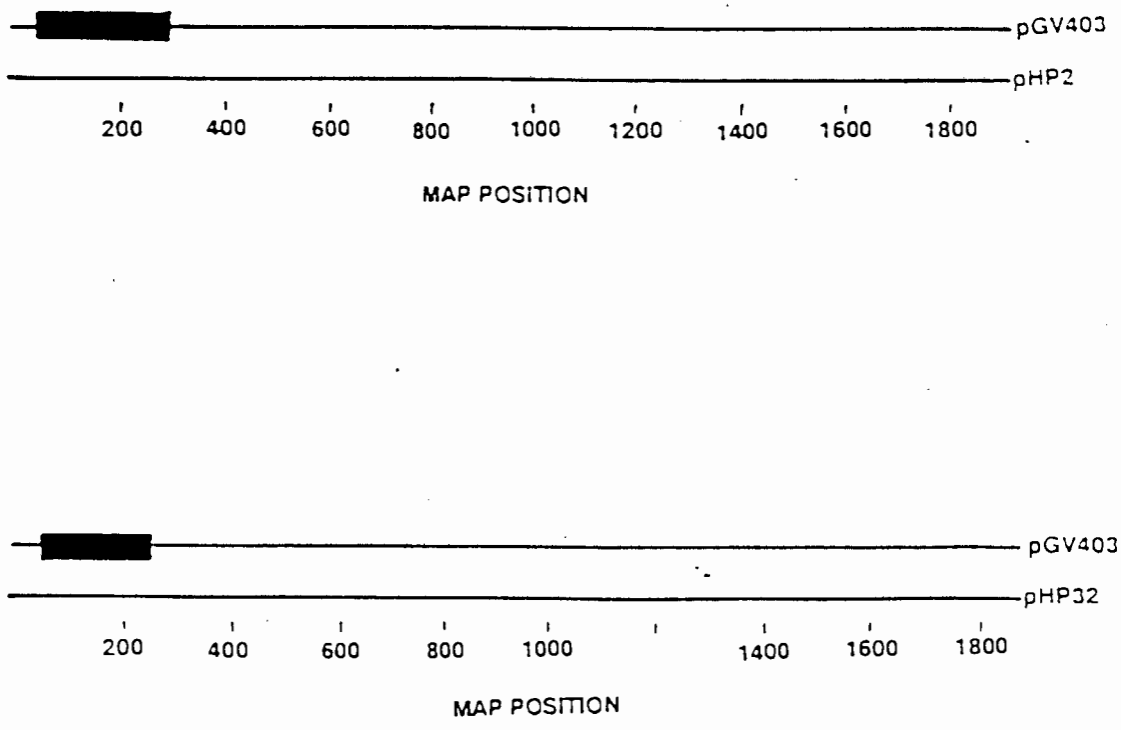


Fig 5. Sequence collineation of **pGV403** with **pHP2/pHP32**
The boxed area in **pGV403** refers to the area inserted in the other plasmids

A similar set of results is observed for **pUC9** and its derivatives.

In this case, however, the transition can not be measured as accurately as before because of the longeroverall sequences involved.

But as shown above there are characteristic transition "peaks"

which are clearly observed in all three plasmid melting profiles (vide supra Fig 4a-c). This holds particularly for the peak numbers II, IV and VI.

Peak number II in the melting profile of **pUC9** (t_m 78.3°C, $\Delta H=5.3*10^3$ kJ/mol) corresponds to peak number II in the profile of **pBR322** (t_m 78.1°C, $\Delta H=5.0*10^3$ kJ/mol) and the corresponding subtransition in the profile of **pBS** plasmid (t_m 78.9°C and $\Delta H=4.6*10^3$ kJ/mol).

A close correspondence between the following peaks is obvious:
Peak number IV of the melting profile of **pUC9** (t_m 82.9°C, $\Delta H=1.4*10^4$ kJ/mol) corresponds to peak number IV of **pBS** (t_m 83.3°C, $\Delta H=1.0*10^4$) and peak number V of **pBR322** (t_m 82.7°C, $\Delta H=1.0*10^4$ kJ/mol).

As a third example the correspondence between the following three peaks is listed below.
Peak number VI of **pUC9** plasmid (t_m 86.3°C, $\Delta H=1.2*10^4$ kJ/mol) corresponds to peak number VI of **pBS** (t_m 87.3°C, $\Delta H=1.4*10^4$ kj/mol) and to peak number IX of **pBR322** plasmid (t_m 87.8°C, $\Delta H=2.2*10^4$ kJ/mol).

It cannot be ruled out that there is a kind of "crosstalk" between neighbouring sequences. Thus, observed differences in the enthalpy and the melting temperature are most probably owing to

slightly different environments i.e. base composition surrounding cooperative melting regions.

Table 5.

A comparison of corresponding transitions measured by DSC in **pGV403** and its derivatives

	pGV403	pHP2	pHP32
peak no	I	I	I
t_m[°C]	79.4	80.0	79.6
ΔH_{CAL} [kJ/mol]	$3.6*10^3$	$3.8*10^3$	$3.3*10^3$
peak no	III	IV	IV
t_m[°C]	83.3	83.2	83.4
ΔH_{CAL} [kJ/mol]	$10.2*10^3$	$12.1*10^3$	$10.0*10^3$

Table 6

A comparison of corresponding transitions measured by ***DSC*** in **pUC9** and its derivatives

	pUC9	pBS	pBR322
peak no	II	II	II
t_m[°C]	78.3	78.1	78.9
ΔH_{CAL} [kJ/mol]	$5.3*10^3$	$4.6*10^3$	$5.0*10^3$
peak no	IV	IV	V
t_m[°C]	82.9	83.3	82.7
ΔH_{CAL} [kJ/mol]	$1.4*10^4$	$1.0*10^4$	$1.0*10^4$
peak no	VI	VI	IX
t_m[°C]	86.3	87.3	87.8
ΔH_{CAL} [kJ/mol]	$1.2*10^4$	$1.4*10^4$	$2.2*10^4$

Summarizing the results obtained by DSC it is possible to conclude that the sequence of any DNA is composed of melting units and that the melting profile is characteristic for a given plasmid.

In recent years a limited number of calorimetric melting curves of selected DNA sequences have been reported[47,53]. Unfortunately none of the published scans have been used for a comparison of the thermal denaturation profile as obtained by DSC and UV-melting curves (thermal denaturation) for the same plasmid or for group of related plasmids. UV-melting curves can be used to calculate ΔH_{VH}. Due to the cooperative nature of the transition ΔH_{VH} can be only interpreted in molecular terms if one can come up with a method which allows one to determine the actual size of the cooperative melting unit.

In the next chapter high resolution thermal melting profiles, as obtained by UV spectroscopy of plasmid DNA are presented and the thermodynamic data are extracted.

CHAPTER III

High resolution thermal denaturation as observed by UV-spectroscopy of plasmid DNA

Introduction

The UV absorption of nucleic acids is due mainly to the electronic states of purine and pyrimidine bases. The other structures, for instance the sugar phosphate backbone, have insignificant contribution to the UV absorption between 200 nm and 300 nm. Nucleotides have very low symmetry and many nonbonded electrons which are responsible for creating several different π-π^* and n-π^* electronic transitions which can be induced by excitations between 200 and 300 nm. The spectra of Guanine and Cytosine show clear evidence for two transitions. On elevating the temperature above a certain threshold value UV absorbance intensity increases about 40% because of release of the bases from their double helical stacking.

It is well established that there exists a linear relationship between melting temperature (t_m) and the gross GC content of a DNA sequence. This has been a very useful tool for the characterization of DNA [4]. It has been shown that thermal denaturation is not a continuous process but can be broken down into a set of discrete steps [5,6,7]. Analysis of each transition (peak) has shown that the GC content is the major factor which determines the observed melting temperature. During melting the (double) DNA helix melts in discrete steps (cooperative melting units);

this is responsible for the creation of internal loops. It was previously shown that short domains, which will create loops of similar or identical base composition are often discernible by small differences in the temperatures at which they dissociate, i.e. their characteristic melting temperatures.

The factors determining the temperature at which domains can exist are: (1) the energy cost of breaking H-bonds and cancelling stacking interactions and (2) the cost of releasing counter ions and the gain in conformational entropy. The methodology of UV-melting was used to study the transition of discrete sequences which were already studied by DSC.

III.1. Thermal denaturation of the linearized plasmid pGV403 and its derivatives

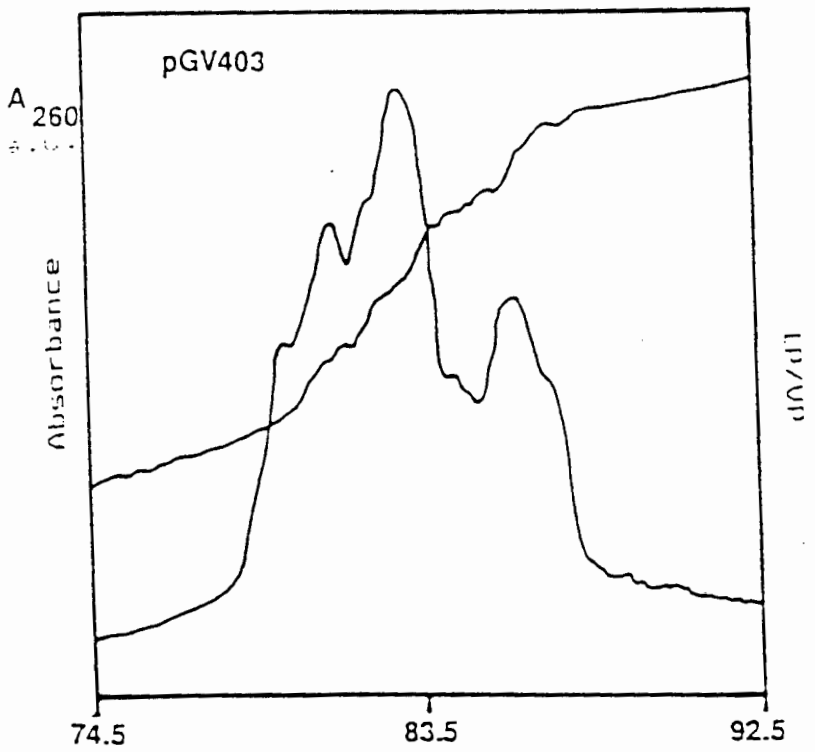
---

All plasmids were linearized prior to denaturation by specific (single-site) restriction endonucleases.

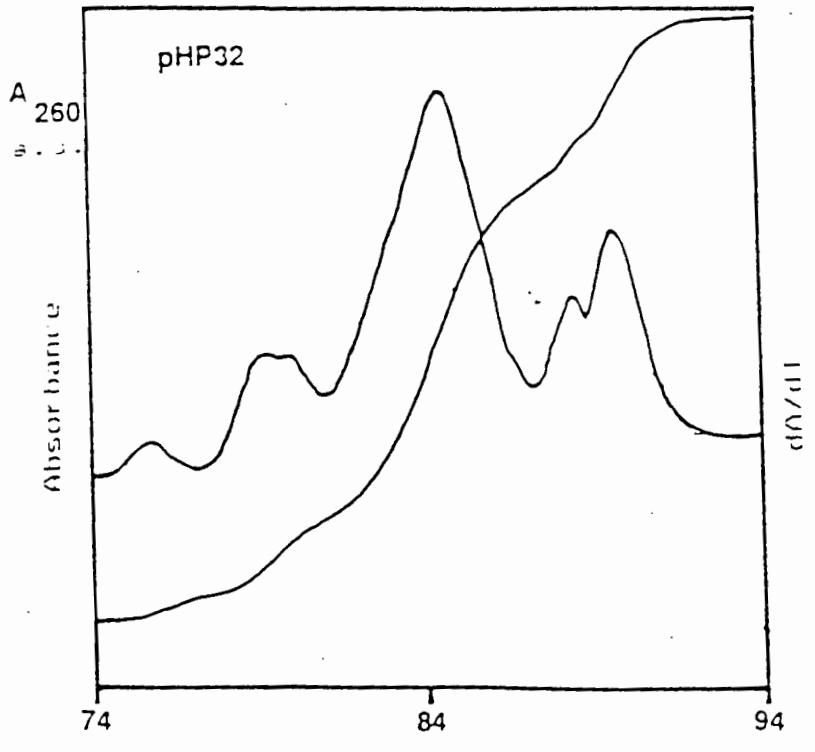
Thermal denaturation of **pGV403** DNA and derivatives (**pHP2** and **pHP32**) was monitored at 260nm (heating rate of 1°C per minute) over the temperature range from 38°C to 98°C in the standard buffer (50mM NaCl, 1mM Tris, 0.5mM EDTA pH 7.6).

As is shown in Figs. 6a-c in the DNA of the plasmid **pGV403** (SmaI linearized) there are seven separable transitions with the following transition temperatures 77.5, 79.1, 80.4, 82.0, 84.2, 86.6, and 88.3 °C. Melting of DNAs from the plasmids **pHP32** and **pHP2** was carried out over the temperature range 68°C to 98°C under the same experimental conditions. Plasmid **pHP2** (PvuII linearized) revealed six transitions with transition temperatures 77.1, 79.6, 80.9, 82.5, 88.0, 89.6°C, and **pHP32** DNA (PvuII linearized) showed six transitions at t_ms of 76.7, 79.2, 80.1, 83.4, 88.4, and 89.6°C

Fig 6a


pGV403
A 260
Absorbance
dA/dT
74.5
83.5
92.5
Temperature °C

6b


pHP32
A 260
dA/dT
74
84
94
Temperature °C

6c

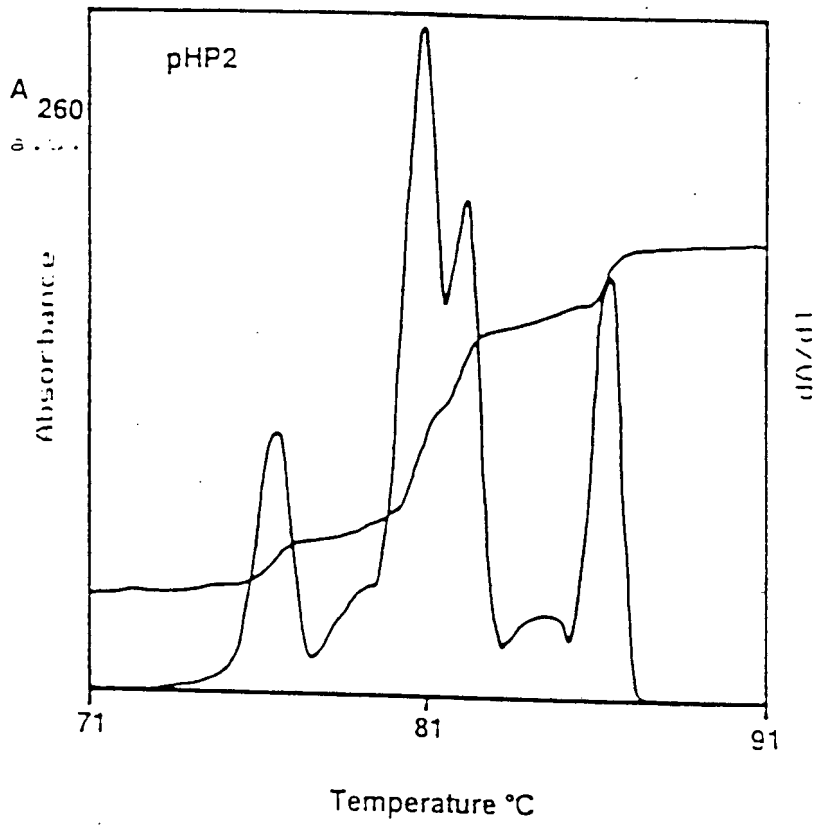


Fig 6.Thermal denaturation profiles and their first derivatives (dA_{260}/dT) versus temperature of plasmids
a) **pGV403** b) **pHP32** c) **pHP2**

III.1.a. Calculation of thermodynamic data for the thermal denaturation of the plasmids pGV403, pHP2 and pHP32

The van't Hoff enthalpy of denaturation ΔH_{VH} was calculated using the standard thermodynamic equation

$$\Delta H_{VH} = (2n+2)\, RT^2 (d\alpha/dT)_{tm} \quad [50]$$

where n=2 for a two stranded helix.

$\Delta H_{VH} = 6*8.31*(77.5)^2*(1/0.136)$ for first transition of plasmid **pGV403**. Table 7 presents the thermodynamic data calculated for the discrete subtransitions of this group of plasmids.

Table 7

	pGV403 ΔH_{VH} [kJ/mol] $*10^3$	pHP2 ΔH_{VH} [kJ/mol] $*10^3$	pHP32 ΔH_{VH} [kJ/mol] $*10^3$
1	2.20	2.28	2.55
2	3.10	5.78	1.51
3	4.22	3.88	3.00
4	5.80	5.86	5.67
5	1.70	3.85	1.00
6	2.91	6.01	2.05
7	1.80		2.43

III.2. Thermal denaturation of the linearized plasmid pUC9 and its derivatives

---

The thermal denaturation profiles for **pUC9** DNA and its derivatives were scanned under the same heating conditions that have been used for the **pGV403** denaturation.

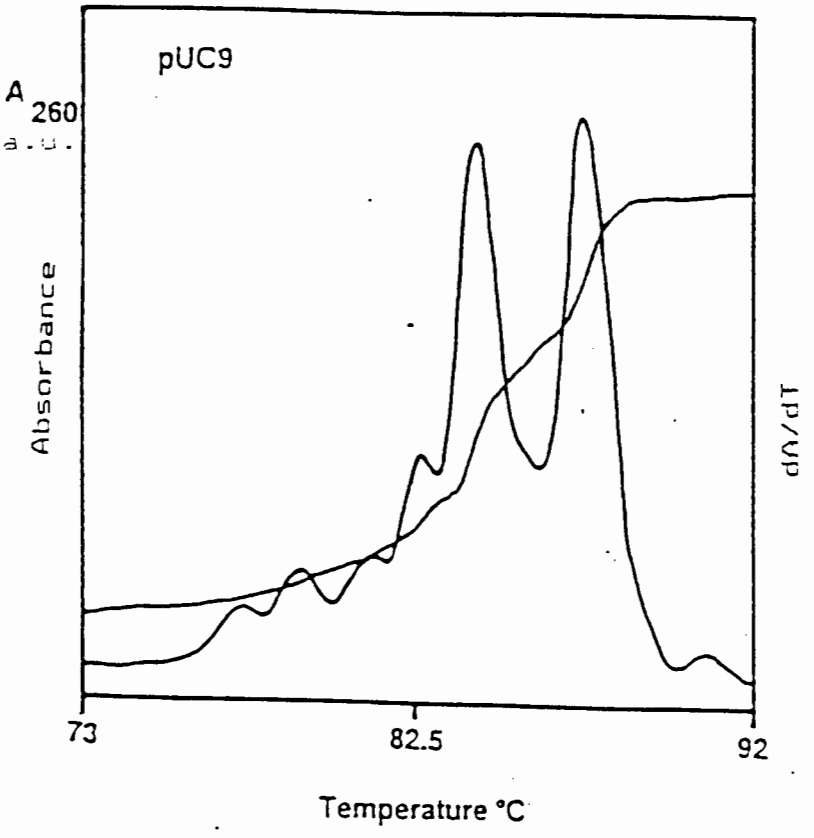
Figs. 7a-c present the melting profiles of **pUC9** DNA, **pBS** DNA and **pBR322** DNA (linearized by EcoRI).

Denaturation of **pUC9** DNA showed 8 transitions at transition temperatures of 76.6, 78.4, 80.8, 82.4, 83.8, 85.8, 87.4, and 89.0°C.

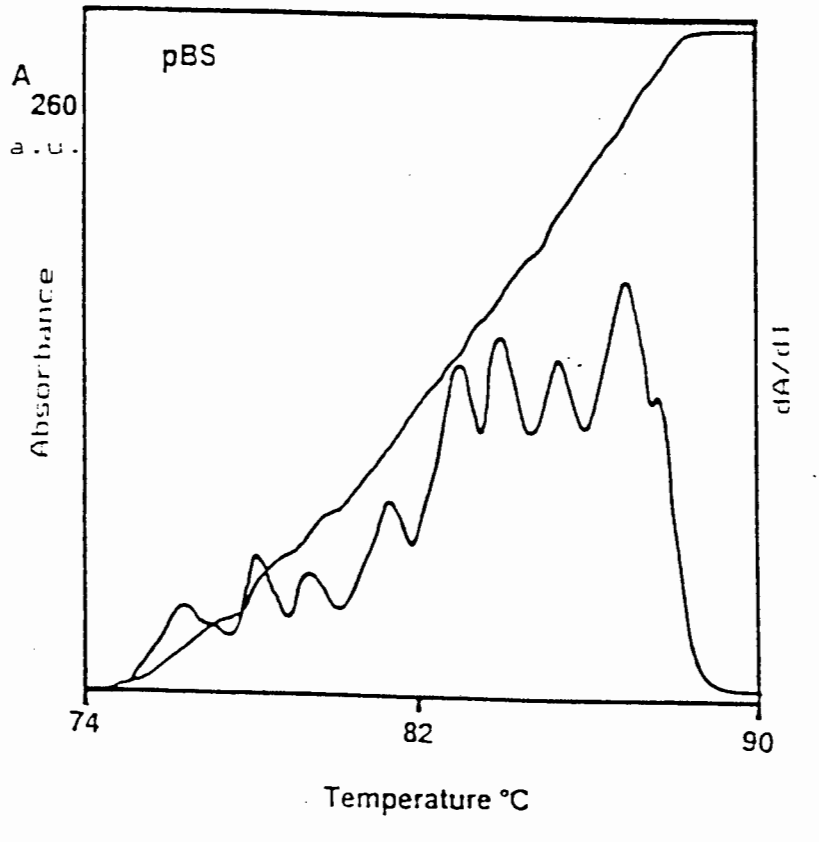
pBR322 DNA is characterized by 9 transitions at transition temperatures of 73.4, 77.0, 78.0, 80.4, 82.2, 84.4, 85.8, 87.2, and 88.2°C.

Finally **pBS** DNA melting can be resolved into 9 transitions at t_ms 76.4, 78.1, 79.6, 81.4, 83.7, 84.8, 86.1, 88.2, and 89.2°C.

Fig 7a


pUC9
A 260 a.u.
Absorbance
dA/dT
73
82.5
92
Temperature °C

7b


pBS
A 260 a.u.
Absorbance
dA/dT
74
82
90
Temperature °C

7c

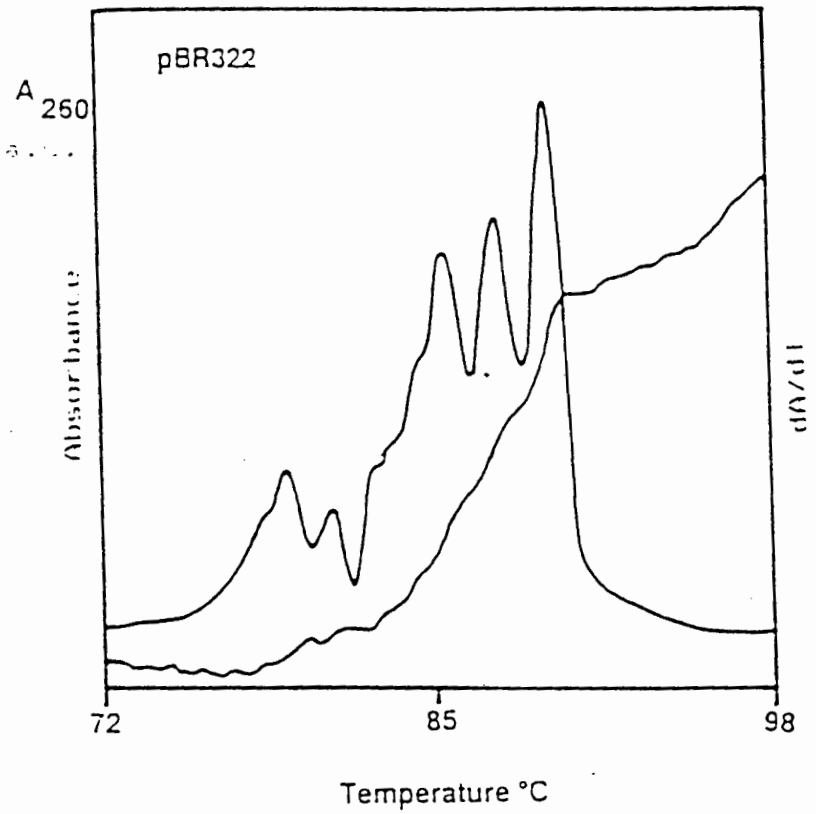


Fig 7. Thermal denaturation profiles and their first derivatives (dA_{260}/dT) versus temperature of plasmids
a) **pUC9** b) **pBS** c) **pBR322**

III.2.a. Calculation of thermodynamic data for the thermal denaturation of the plasmids pUC9, pBS and pBR322

For each subtransition the van't Hoff enthalpy can be calculated as shown in III.1.a. The data obtained for individual transitions characteristic of the group of plasmids is shown in Table 8.

Table 8.

	pUC9 ΔH_{VH} [kJ/mol] $*10^3$	pBS ΔH_{VH} [kJ/mol] $*10^3$	pBR322 ΔH_{VH} [kJ/mol] $*10^3$
1	2.02	5.21	3.70
2	4.00	4.84	4.26
3	1.52	5.89	7.51
4	3.93	3.57	2.38
5	7.92	6.03	5.12
6	5.34	4.96	2.22
7	4.62	11.00	5.27
8	8.15	13.92	3.11
9		8.00	3.97

Summary and conclusion of the high resolution thermal denaturation, observed by UV-spectroscopy

---

The results presented in this chapter show that there are characteristic features in thermal stability profiles (i.e. DNA structure) within this group of plasmids which can be attributed to corresponding sections in their sequences (similar melting domains). To demonstrate this in more detail the following examples are selected: Looking at the melting profile of **pGV403** one can see that transition (peak) number I corresponds closely to the first transition in the melting profile of **pHP2** and to peak number II of **pHP32**. Similar correspondences hold e.g. for peak number IV of **pGV403**. This transition corresponds to peak number IV of **pHP2** and peak number IV of **pHP32.** A similar set of results was obtained for **pUC9** and its derivatives. Peak number II of **pUC9** plasmid corresponds well to peak number II of **pBS** and the peak number II of **pBR322** respectively.

Peak number IV of **pUC9** corresponds to the peak number IV of **pBS** and peak number IV of **pBR322**, and finally peak number VI of **pUC9** corresponds to the peak number VI of **pBS** and peak number VII of **pBR322**.

The correspondence between the peaks is based upon the transition temperatures (t_m) and the transition enthalpy (ΔH_{VH}).

Over and above the correspondence of certain peaks there appears to be a general similarity within melting profiles of the two groups, although, as expected, the melting profile of **pBR322** has a more complicated

pattern because of the larger size of the sequence and therefore it will have a more complex base sequence structure.

The results discussed above are compiled in Tables 9 and 10.

Table 9

Transition temperatures and transition enthalpies of selected peaks

	pGV403	pHP2	pHP32
peak no	I	I	I
temperature[°C]	77.5	77.1	76.7
ΔH_{VH} [kJ/mol]*10^3	2.20	2.28	2.55
peak no	IV	IV	IV
temperature[°C]	82.0	82.5	83.4
ΔH_{VH} [kJ/mol]*10^3	5.80	5.86	5.67

Table 10.

Transition temperatures and transition enthalpies of selected peaks

	pUC9	pBS	pBR322
peak no	II	II	II
temperature[°C]	78.4	78.1	78.0
ΔH_{VH} [kJ/mol]*10^3	4.95	5.84	5.00
peak no	IV	IV	IV
temperature[°C]	82.4	81.4	82.2
ΔH_{VH} [kJ/mol]*10^3	3.93	3.55	3.12
peak no	VI	VI	VII
temperature[°C]	85.8	84.8	85.8
ΔH_{VH} [kJ/mol]*10^3	5.34	4.96	5.27

From the results discussed in this chapter and in the previous chapter it is evident that there are remarkable similarities between the data obtained from DSC and from UV-thermal denaturation studies. The **pGV403** family of plasmids is characterized by 6 or 7 discrete transitions which can be seen by either method, whereas the second group of plasmids, i.e. the **pUC9** family, exhibits 8 or 9 transitions using either one of the techniques. The similarities rest upon the temperatures of the

transitions (t_m), detectable by either technique. It is possible that the differences observed for t_ms of corresponding transitions could be either due to (1) variation in accuracy between the techniques and/or (2) artefacts introduced by the linearization of the plasmids for the thermal denaturation studies. There is a slight possibility that one observes a different transition if plasmids are linearized with other (single-site) restriction endonucleases. The results obtained by these two techniques clearly support each other and show that different regions of DNA sequences undergo denaturation at very distinct and well defined temperatures. Owing to the close resemblance of the results it is reasonable to assume that the same sequences of the DNA are melting and give rise to the corresponding results.

In the next chapter an attempt will be made to identify the location of the cooperative melting regions by a semiempirical as well as by a direct experimental approach. The semiempirical approach will be based on the loop entropy function. The experimental approach tries to visualize the denatured sequences with the help of E.M.

CHAPTER IV

Loop energy calculation based on plasmid DNA denaturation

Introduction

In order to extend the approach to establish a correlation between the gross GC content in various sequence domains and the transition temperature a semiempirical calculation was adapted which was introduced by Blake *et al* [49,54]. The equation according to which F_{GC} (*vide infra*) was determined is given below. The raw data are taken from the high resolution UV-melting profile [29,49,54].

As the experimental results indicate, the melting of DNA regions with different thermodynamic stabilities occurs in a stepwise fashion [40,42]. This regional melting leads inevitably to the creation of internal loops. Each loop has its own internal energy which is reflected in ***DSC*** scans and UV-melting profiles by the appearance of discrete transitions (peaks) at well defined temperatures (t_ms).

F_{GC} was calculated accordingly to the following equation:

$$\mathbf{F_{GC} = [\ 3.059B-0.204]/[2.350B+1.493]}$$

where $\mathbf{B= \Delta A_{282}/\Delta A_{260}}$

F_{GC} [49,54] refers to the fraction of all GC base pairs in the total genome which are denatured at any given temperature. The two wavelength recording capability of the **Hewlett Packard 8450A Diode Array Spectrophotometer** was used to study the melting of two different families of plasmids. Both have a wide range of base composition distributed over lengths from 1665bp(**pGV403**) to 4363bp (**pBR322**) [30,31,45].

In each case the melting of 25 μg of plasmid (0.5 OD) in standard buffer was followed simultaneously at 260 and 282 nm over the temperature range 68°C-98°C.

All experiments were carried out under a set of standard conditions: 50mM NaCl, 0.5mM EDTA, 1 mM TRIS pH 7.6 at the heating rate of 1°C per minute and the temperature range of 68°C-98°C

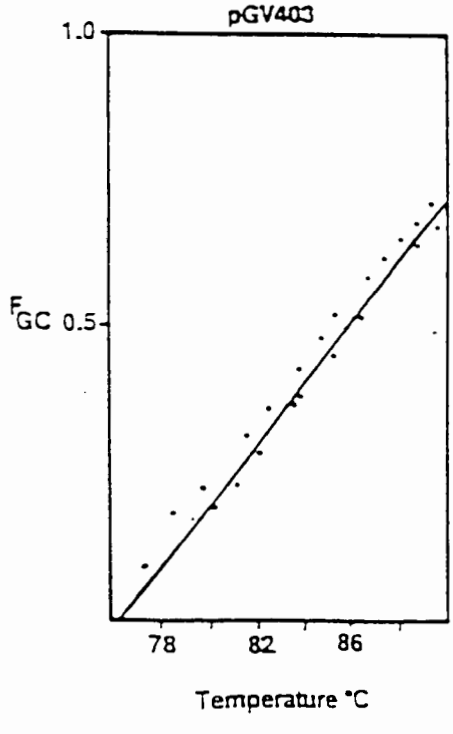
IV.1 Determination of F_{GC} in the plasmid pGV403 and its derivatives

---

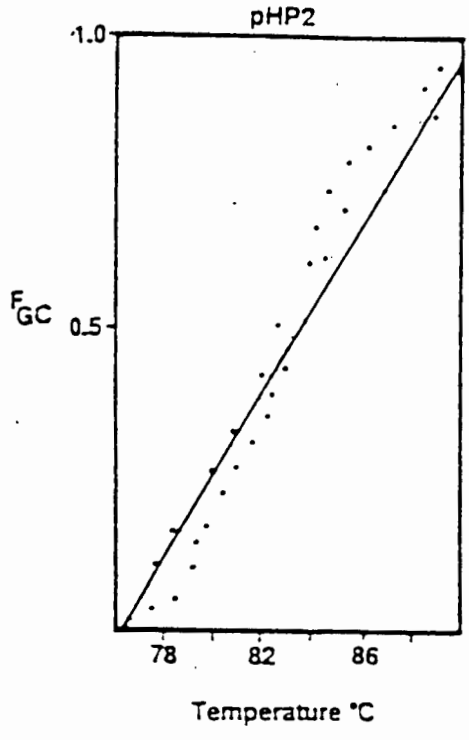
The ratio $\Delta A_{282}/\Delta A_{260}$ was monitored over the temperature range from 68 to 98°C with a heating rate of 1°C per minute. Data acquisition results in 20 absorbance readings per minute. The data were compiled in an ASCII file and processed with the help of the HP computer in connection with **Hewlett Packard 8450A Diode Array Spectrophotometer**.

In each experiment, 25 µg of plasmid was used.
The results obtained for the melting of linearized **pGV403** DNA and its derivatives are shown in Figs 8a-c.

Fig 8a



8b



8c

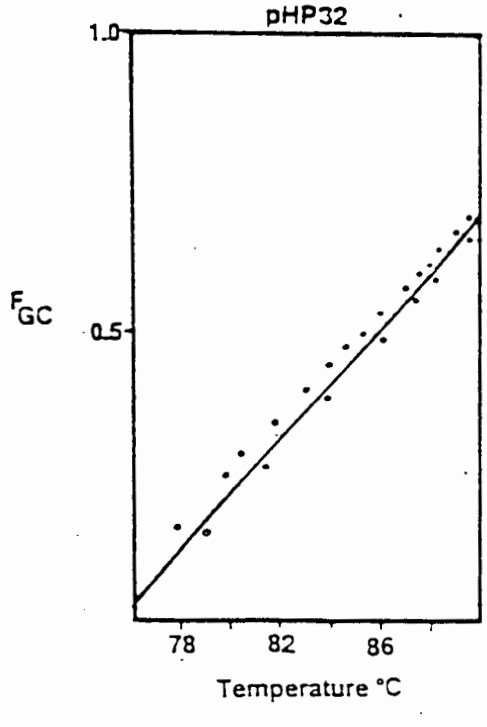


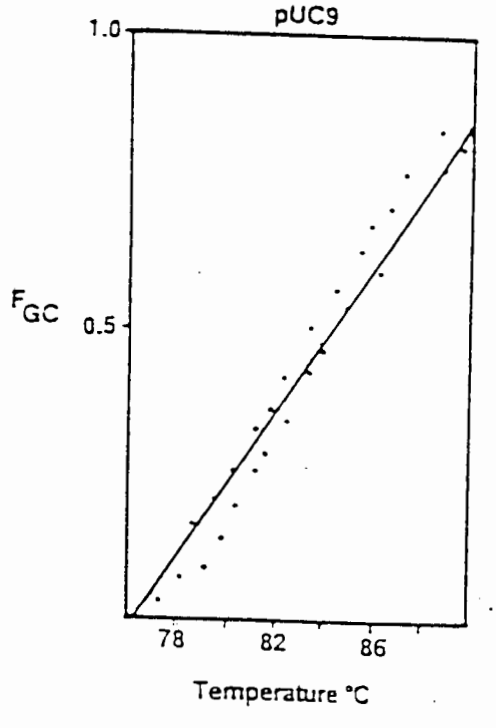
Fig 8. The increase of F_{GC} with temperature for the linearized plasmids a) **pGV403** b) **pHP2** c)**pHP32.** The points are obtained according to the equation given above.
The solid lines were obtained by least square regression analysis.

IV.2. Determination of F_{GC} in the plasmid pUC9 and its derivatives

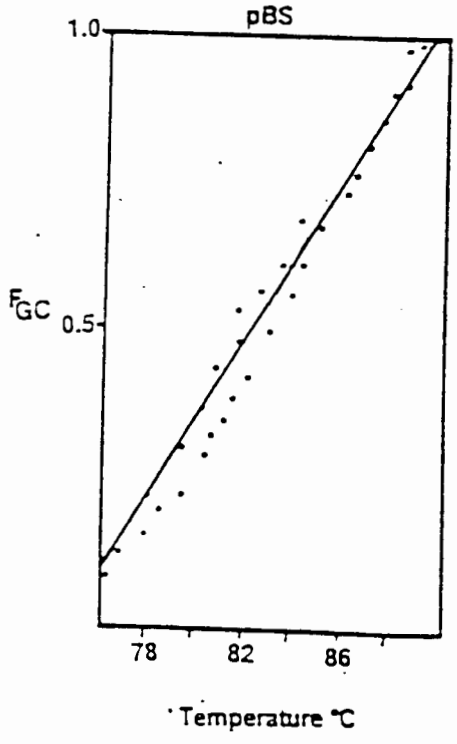
The high resolution melting data at A_{282} and A_{260} of the **pUC9** family of plasmid sequences were treated in the same way as for the **pGV403** family.

The results are shown in Fig 9a-c.

Fig 9a



9b



9c

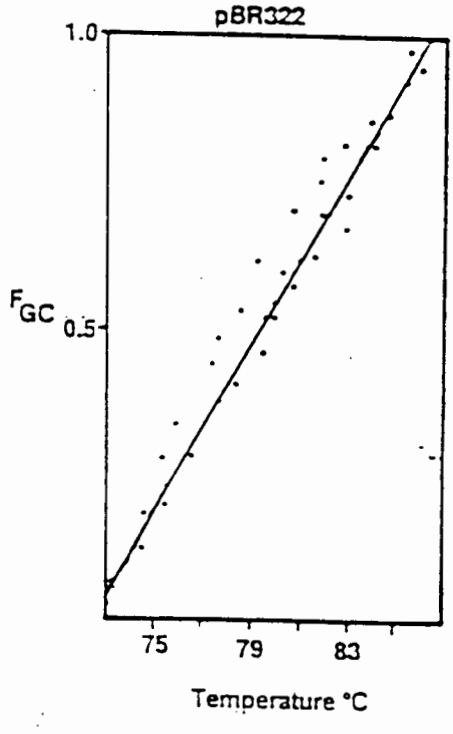


Fig 9. The increase of F_{GC} with temperature for the linearized plasmids a) **pUC9** b) **pBS** c)**pBR322.** The points are obtained according to the equation given above.
The solid lines were obtained by least square regression analysis

Summary and conclusion of loop energy calculation

It can be seen from the results shown in Figs 8 and 9 that the base composition of dissociating domains in all plasmids increases linearly with their relative GC content. The results show clearly that the stability of internal loops depends only on their GC content and therefore the stacking energy of these base-pairs.

The results obtained are in good agreement with theoretical and experimental results obtained previously [54].

CHAPTER V

Complexity as a means to determine functional domain in plasmid DNA

Introduction

Following the suggestion of Schrödinger [1] that life processes feed on "neg-entropy" the common approach is based on the assumption that biological entropies are of pronounced order. To compare different systems one needs a objective measure for the degree of order. The objective scale for the degree of order is the entropy of the system. Shannon [2] was the first to point out that there is a correlation between entropy and information, and Szillard based his explanation of the Maxwell demon on this correlation. The demon can only discriminate between fast and slow particles on the basis of information. The amount of information needed for effective decision making is equal to the entropy of the system. It is known that "normal" DNA consists of four different bases which can be classified into two groups, namely purines and pyrimidines. Considering these bases as letters of the four letter alphabet (similar to dots and dashes in the telegraphic code) i.e. Arginine-A, Guanosine-G, Cytosine-C and Thymine-T respectively. One can presume that within the DNA sequence information can be encoded. A living cell can be considered as an information system with an information matrix which is the DNA molecule. Following the information system's classification of Shanon[2], DNA can be treated as a discrete system where the output signal consists of a set of discrete signals (A, T,

G, C). For some time it has been known that within a DNA molecule regions exist with defined biochemical and physical properties and the most important among them are called genes. Knowing that genes are responsible for the synthesis of proteins with defined physical and biological properties, one can presume that information encoded within the base sequence is not an entirely random feature of the DNA. Thus DNA may be treated as a "language" with "words" written with four different letters. Considering the above one can ask the following questions: Is the probability of the occurrence of the chosen letters i.e. A, T, G, C along the DNA sequence entirely random and how is it measured? As an indicator of randomness of the distribution of bases within the chosen sequence, the local entropy, given by the equation below, was defined.

$$\mathbf{S = 1/L*[A*logb(A/L) + G*logb(G/L) + T*logb(T/L) + C*logb(C/L)]}$$

where L is the length of the sliding window (in the number of bases)
b is the base of the logarithm
According to this equation the local entropy (**S**) is equal to 0 for a homopolymer and approaches a maximum for entirely random sequences (probability of each base within the chosen sequence is equal to 25%).

Because there was no procedure to calculate the local compositional complexity it was necessary to develop one. The procedure is presented in Appendix C.
Calculations were carried out on two groups of plasmids: **pGV403**, **pHP2**, **pHP32** and **pUC9**, **pBS** and **pBR322**.

V.1 Local Compositional Complexity of the plasmid pGV403 and its derivatives

---

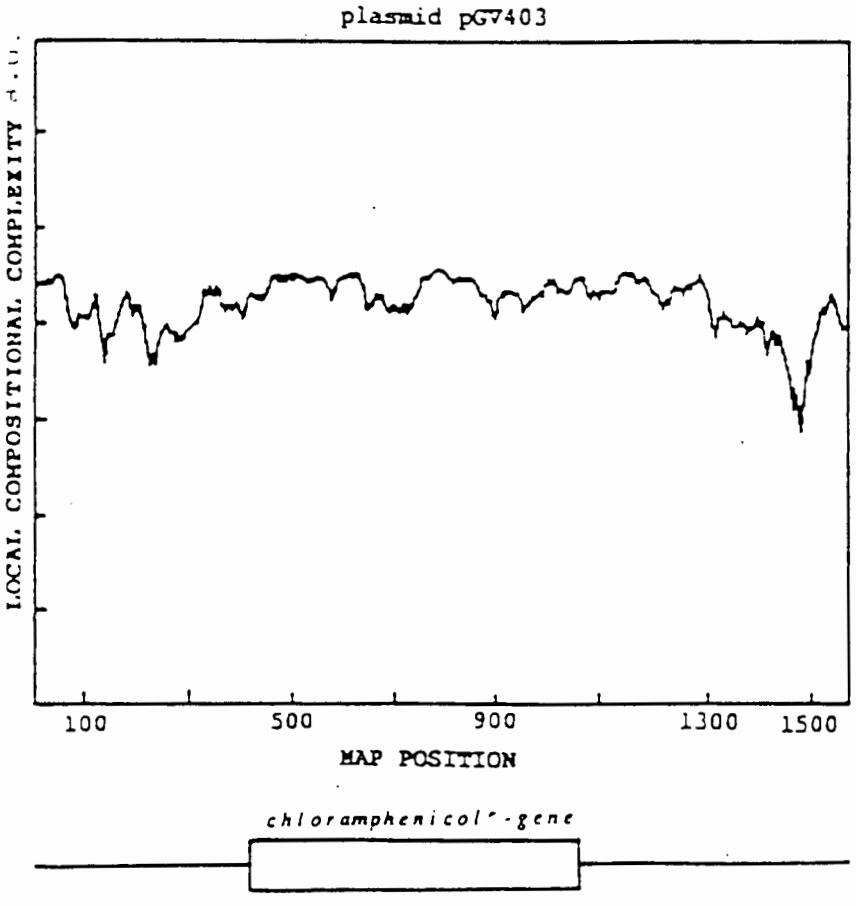
Compositional complexity was evaluated by sliding a 100 bp wide window along the sequence of the plasmid in steps of 1 bp width. The local entropy (cf p 67) which is a measure of the complexity was plotted against the base position at the middle of the window. The calculated complexity is clearly distinguishable from the results one would obtain from a completely random nucleotide sequence. The local complexity minima reflect sequences of particular order.

As can be seen from Figs 10a-c all sequences within this group of plasmids reveal a similar complexity pattern. At the beginning of the **pHP2** and **pHP32** sequences respectively a particular sequence was inserted into the parental plasmid **pGV403**. The nucleotide sequences of these inserts are presented in Figs 11a and b.

The complexity profiles for the plasmids **pHP2** and **pHP32** differ markedly at these early map positions from the parental (**pGV403**) sequence. This marked difference is clearly a result of the insertion of the extra sequences into these plasmids. This is the first time that a formal method is presented to detect and scale these differences and relate them to the information content. Specifically at position 150 in the plasmid **pHP2** and **pHP32** a new "peak" appeared, which strongly suggests that there is an insertion of "foreign" DNA into the parental sequence of **pGV403**. Apart from this change, the profile is quite conserved within this plasmid group. The Chloramphenicol resistance gene can be characterized by a "plateau" (a section of low local entropy variation),

which has a similar shape in all plasmids in this group. This result is supported by Wada's results on the melting of complex genomes where he finds that genes are homeostable sequences [51].

Fig 10a



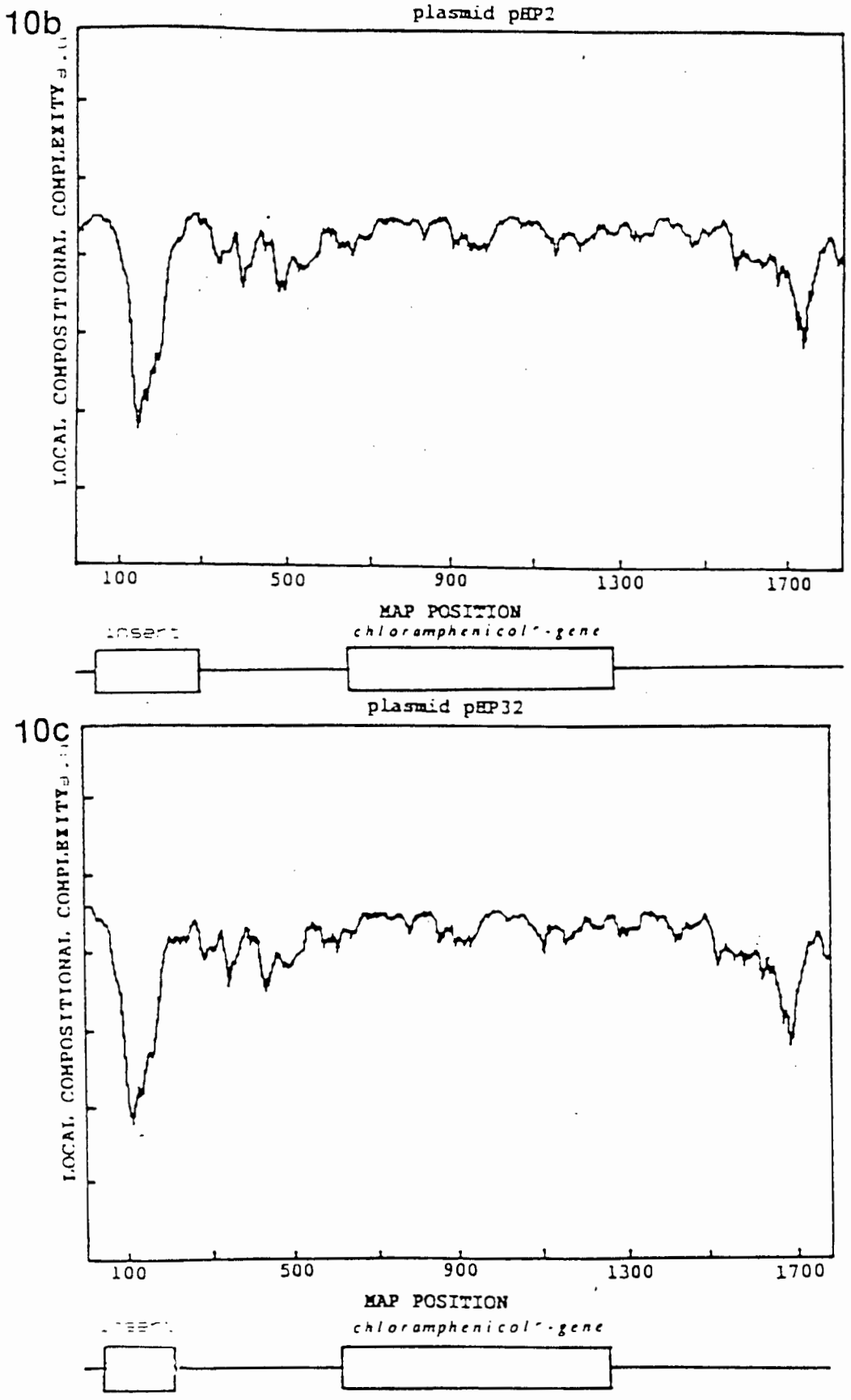


Fig 10. Local compositional complexity versus map position
a) **pGV403** b) **pHP2** c)**pHP32**
All calculations were evaluated using a sliding 100bp window, with 1bp steps along the sequence.
A simplified linear map of the plasmids is also shown.

Fig 11a

```
  51    TCGAATATTAGTAGGAATATGTGCGCGTCAGCTAGTCTACTTTTCAAG 100
 101 TAATTGCGATGTAAATGTCACAAAACCCGTTAAGAGGGAGGGGGGGGGGG 150
 151 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGGGAAGGAGATTTATACAA 200
 201 CTTAACTCACACAAAGTACAGTCCTCACCATATTGAGAGAAAAGGAGATG 250
 251 CAC
```

11b

```
  51    AGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCCCAGCTTATAATC 100
 101 ATCCTTATACACGCGCAGTCGATGAGATGAAAAGTTCATTAACGCTACAT 150
 151 TTACAGTGTTTTGGGCAATTCTCCCTCCCCCCCCCCCTCTCTCTCTCTCT 200
 201 CTCTCTCTCTCTCTCTCTCCCTTCCTCTAAATATGTTGAATTGAGTGTGT 250
 251 TTCATGTCAGGAGTGGTATAACTCTCTTTTCCTCTACGTGGGGTACCGAG 300
 301 CTCGAATT
```

Fig 11. Sequences of DNA introduced into **pGV403** to generate a) **pHP32** b) **pHP2** plasmids.

V.2 Local Compositional Complexity of the plasmid pUC9 and its derivatives

---

The same sliding window approach was applied to the **pUC9** plasmid family and the local compositional complexity was evaluated using the same sliding window of 100 bp. The results of the local entropy obtained are presented in Figs 12a-c.

The results indicate that the Ampicillin resistance gene showed a fairly stable and uniform profile across the group of plasmids reinforcing the claim that genes are homeostable sequences.

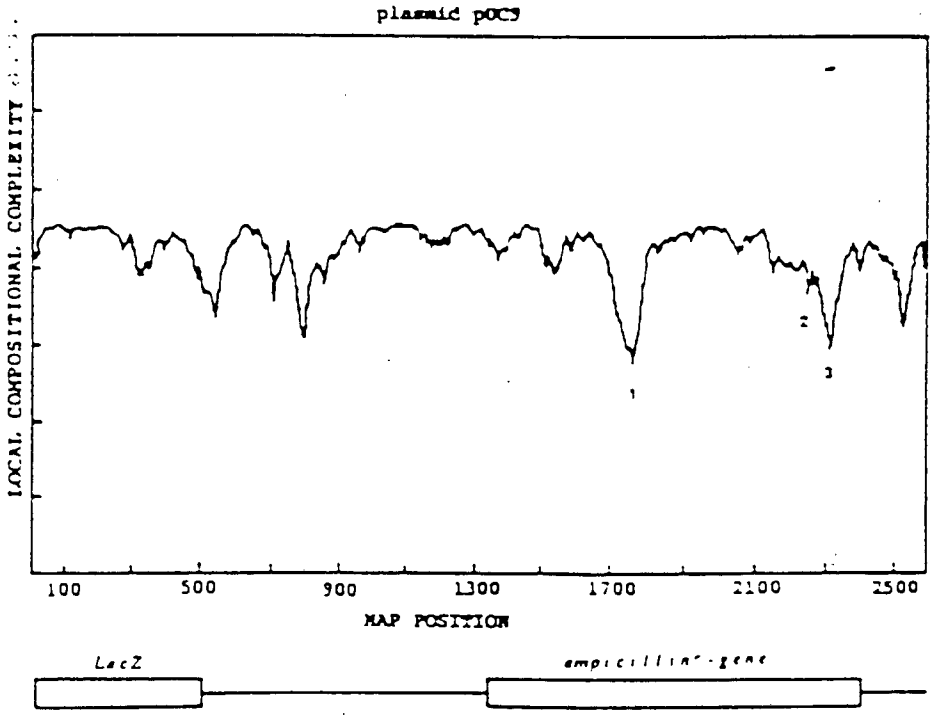
It can be seen from Fig 12a-c that the Ampicillin resistance gene is characterized by three well-defined peaks (numbered 1, 2, and 3) and a similar smooth area between peaks 1 and 2. It is tempting to suggest that the peaks reflect the complexity of regulatory sequences while the smooth area reflects the coding sequences. But at this stage this is only a hint and the rigorous proof for this claim will need further investigation.

The LacZ operon included in **pUC9** and **pBS** also generates a characteristic profile which is absent from the profile of the plasmid **pBR322**.

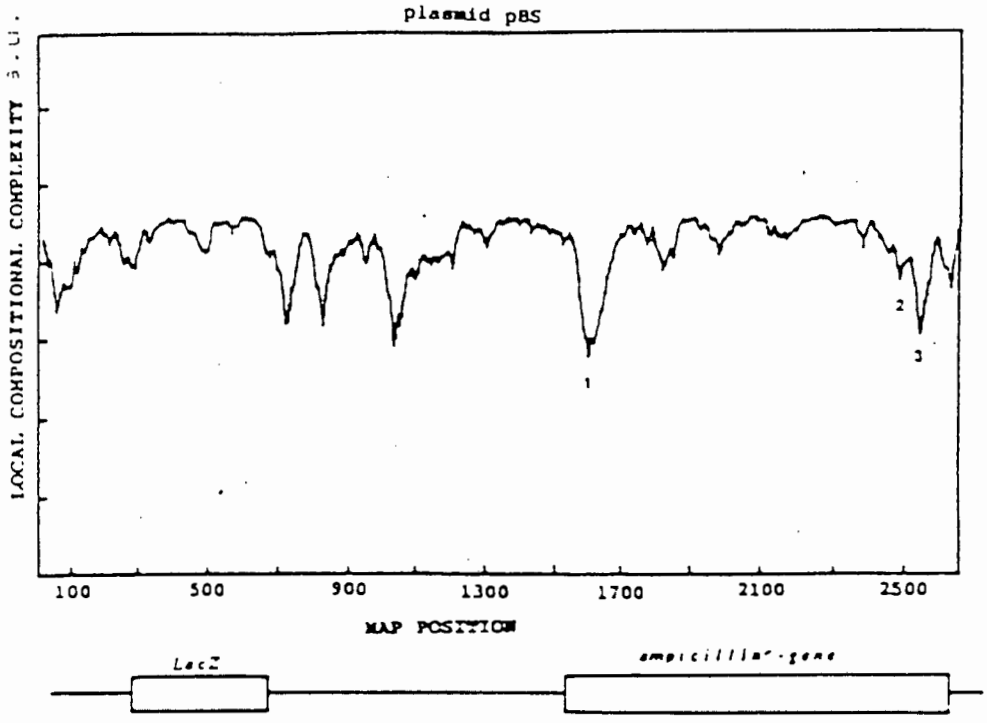
Local compositional complexity exhibits a large variety of peak pattern which, at a later stage, will be a useful tool to analyze genomes and to calculate local thermodynamic properties of sequences, underlining the

claim of this thesis that there is a close correspondence between the physical and the genetic map of a given genome.

Fig 12a



12b



12c

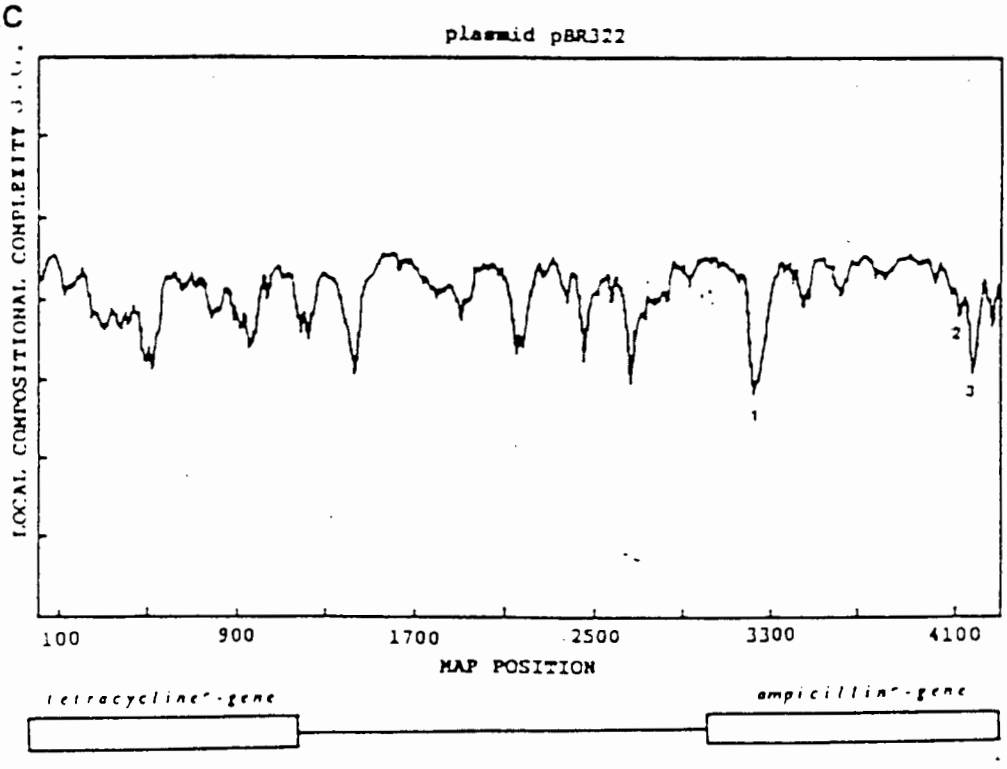


Fig 12. Local compositional complexity versus map position
a) **pUC9** b) **pBS** c) **pBR322**
All calculations were evaluated using a sliding 100bp window, with 1bp steps along the sequence.
A simplified linear map of the plasmids is also shown.

Summary and conclusion of the local complexity analysis of linearized plasmids

---

LCC (local compositional complexity) can be defined as the entropy corresponding to the set of probabilities of the occurrence of the nucleotides G, C, A, and T which are carrying specific information. Based on this fundamental fact the local compositional complexity can be calculated for the four variables, A, G, C, and T respectively. This research concentrated specifically on genes and their contribution to the local compositional complexity of plasmid DNAs, i.e. smooth complexity patterns can be attributed to the coding sequences. The steep gradients of local complexity variations result in peaks, which can be used to fingerprint plasmid families.

As one can see from the procedure presented in Appendix C there were a few reasons for choosing a window width of 100 bp.

1. There is a finite probability that in a smaller window one of the nucleotides may not occur and $\log_b 0$ is equal to $-\infty$.
2. There is an increasing amount of data which suggests that 100 bp is the cooperative melting unit for DNA [40].

The results presented in this chapter clearly support the hypothesis that LCC analysis "reveals" biologically important structures within DNA.

The gene can be described as a region with a complexity which is unperturbed across a range (population) of plasmids, provided that the sequence of the gene is not "damaged", i.e. changed by insertion or deletion of major fragments.

A good example of the influence of inserted (foreign) DNA on LCC was demonstrated for **pGV403** and its derivatives. Assuming that the LCC pattern of **pGV403** is the parental profile, it can readily be seen that insertions can create abrupt changes in the pattern of LCC (Figs 10a-c). The LCC profile of the second group of plasmids i.e. **pUC9** and its derivatives, is not as well defined as the pattern in the first group. One of the reasons for this may be the greater length of the plasmids in this group.

The Ampicillin resistance gene can be described by three "peaks": 1, 2 and 3, a pattern which can serve as a landmark for the whole group of plasmids.

In a study published by Konopka and Owens [57], it was suggested that LCC was lower in exons (of eukaryotic DNA) and higher in introns. This could imply that prokaryotic DNA may be "simpler" in terms of complexity than the eukaryotic genomic DNA. The results presented here for plasmid DNA suggest that this is not necessarily the case, since structural units as shown (genes) for the prokaryotic genome are well defined by local compositional complexity.

CHAPTER VI

Electron microscopy: Visualization of local thermal denaturation of DNA

Introduction

One way to identify partially melted regions of DNA is by using electron microscopy, a method which allows one to trap these regions by "freezing" them through a chemical reaction.

Electron microscopy of melted plasmid DNA has been done in recent years [16,21,39]. This methodology gives information about the peculiarities of the nucleotide distribution in the DNA (melting maps), and together with ***DSC*** and spectrophotometric methods (melting curves) can have important applications in molecular genetics. It provides information regarding the distribution of bases along the DNA and the distribution of genetically and thermodynamically important units. Comparison of EM melting maps and UV/DSC melting curves can be a useful tool for the physical characterization of DNA sequences.

For the fixation of molten regions a 0.1M glyoxal solution was added to the buffered DNA solution at selected temperatures. The glyoxal reacts with guanine to give a very stable reaction product. By this procedure

denatured (molten) regions corresponding to the melting transitions (peaks) up to the given temperature can be "frozen in".

Each individual molecule examined using the electron microscope will show its unique pattern i.e. single stranded loops or sometimes loops collapsed into globular structures interspersed with double stranded sections. The location of loops is recorded by measuring distances from the ends of the molecule. This is repeated for a large number of individual molecules to yield reliable values.

A set of melting maps constructed from electron micrographs made it possible to determine the size and location of the DNA regions corresponding to the transitions (peaks) in the profile of the melting curve. Specific bases in a given position can be located with a precision of +/- 150 Å. Experiments were done on one representative each of the two plasmid families i.e. **pGV403** and **pUC9.**

VI.1. plasmid pGV403

Based on the experience obtained from high resolution thermal denaturation experiments (results not shown) it has become obvious that the linear form of plasmid DNA has a better defined melting profile than an open circular one.

Fig 13 shows the melting curve of **pGV403** relaxed by a single cut of SmaI.

It can be observed that the melting profile has seven clearly distinguishable transitions (peaks). In order to locate various regions on the plasmid and establish which of the peaks actually corresponds to each region, plasmid DNA was fixed with 0.1M glyoxal at selected temperatures as indicated by arrows.

Electron micrographs (Fig 14) were digitized in order to obtain melting maps at each melting stage.

Fig 13

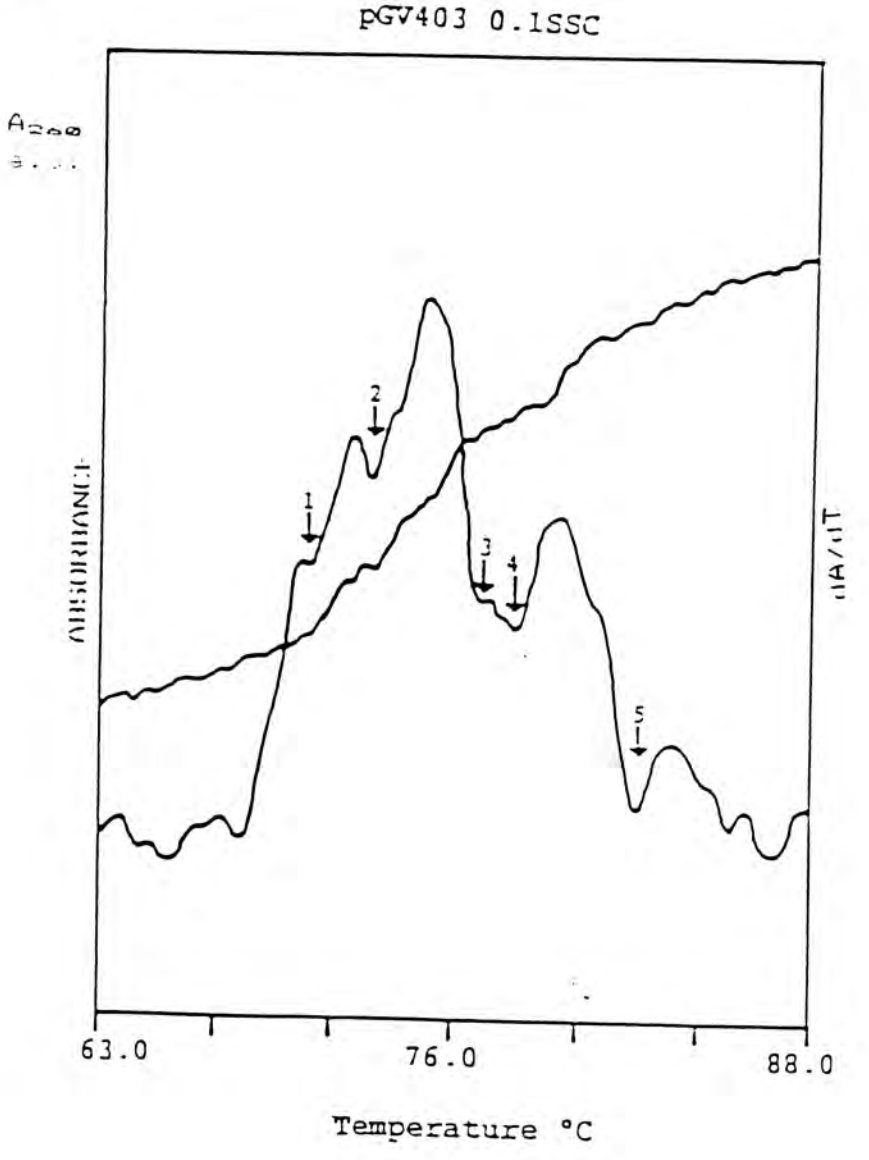


Fig 13. Thermal denaturation curve and first derivative for **pGV403** in 0.1*SSC. Arrows indicate the fixing temperature of partially denatured DNA.
1) 70.8°C 2) 72.2°C 3) 76.4°C 4) 77.8°C 5) 81.5°C

Fig 14

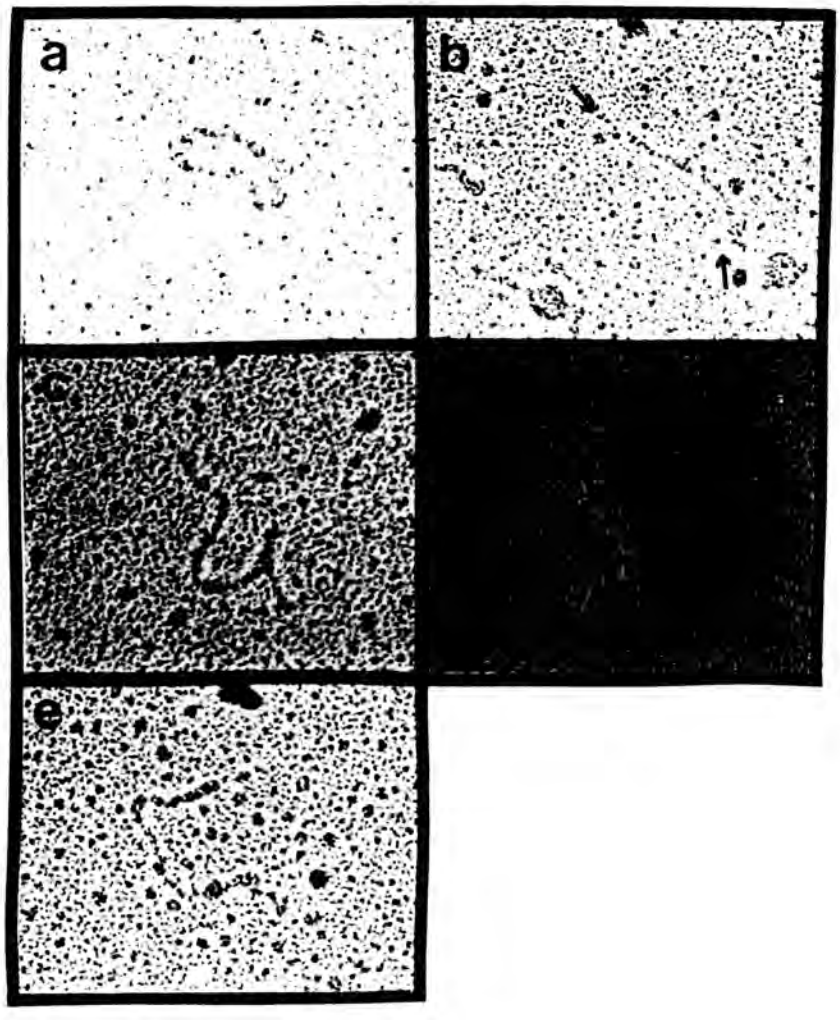
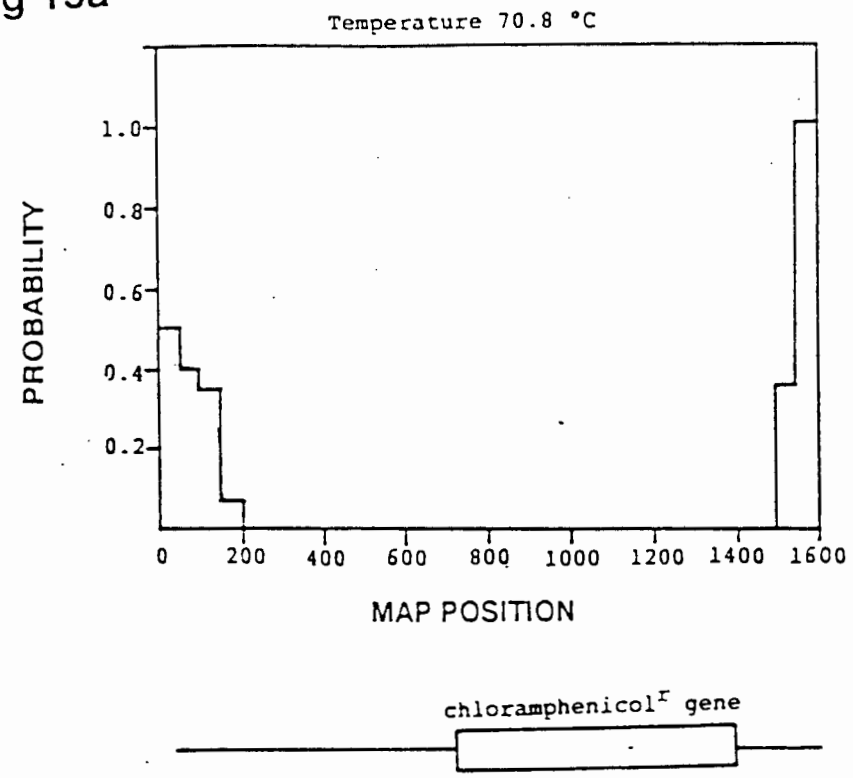
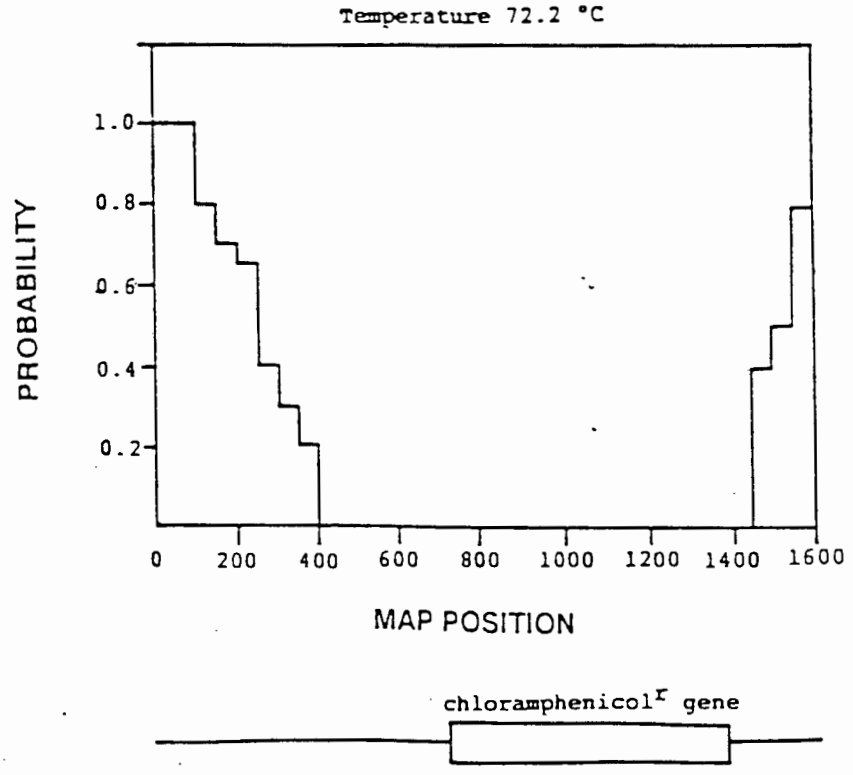


Fig 14. Electron micrographs of partially denatured linearized DNA of **pGV403** plasmid.
a) circular form 25°C b) 70.8°C c) 72.2°C d) 76.4°C
d) 77.8°C

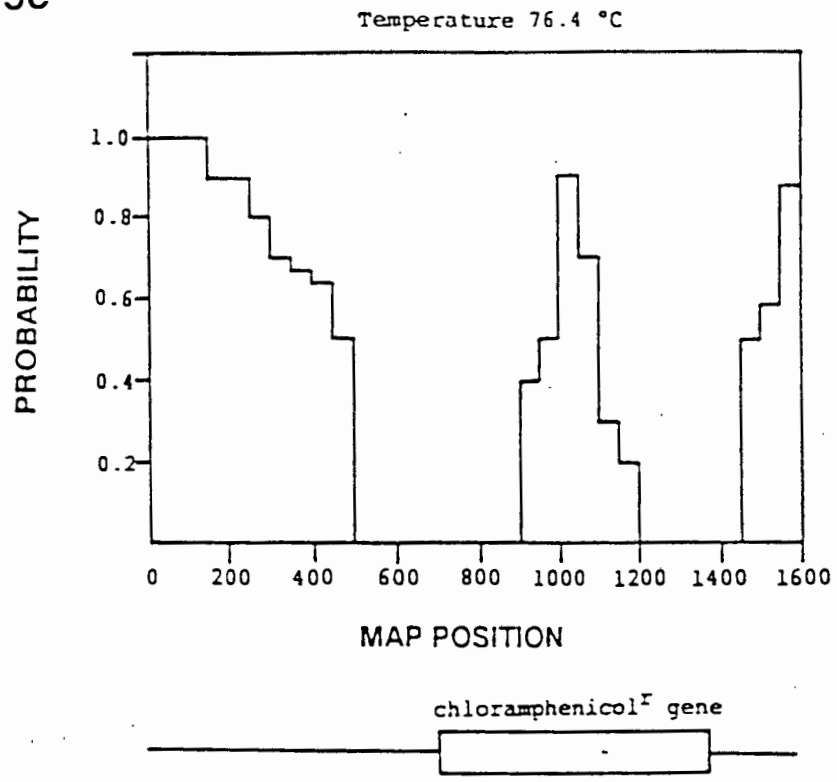
Fig 15a


Temperature 70.8 °C
PROBABILITY
1.0
0.8
0.6
0.4
0.2
0
200
400
600
800
1000
1200
1400
1600
MAP POSITION
chloramphenicol^{r} gene

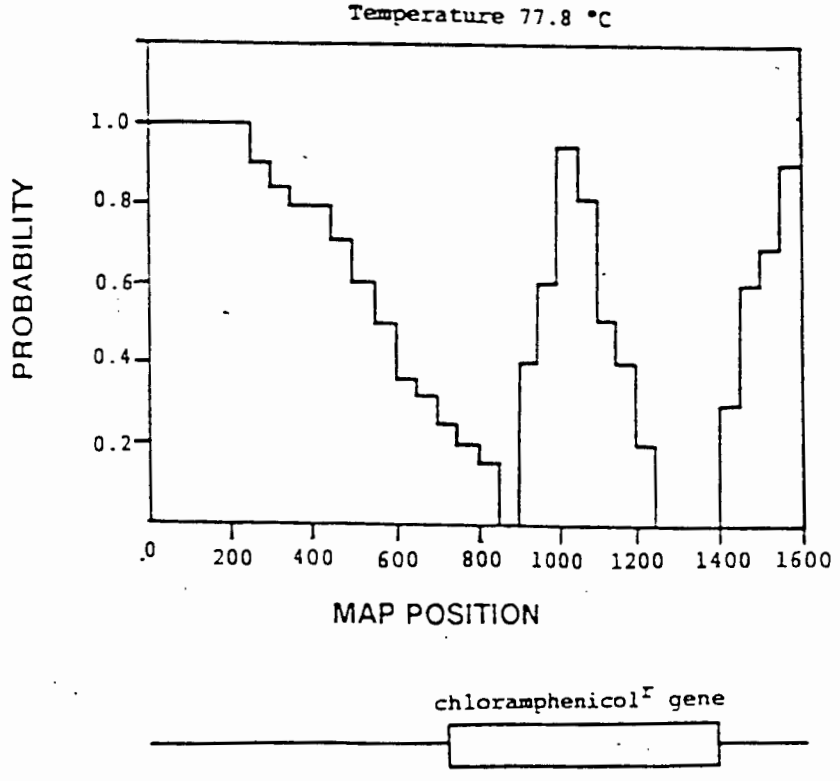
15b


Temperature 72.2 °C
PROBABILITY
1.0
0.8
0.6
0.4
0.2
0
200
400
600
800
1000
1200
1400
1600
MAP POSITION
chloramphenicol^{r} gene

15c


Temperature 76.4 °C
PROBABILITY
1.0
0.8
0.6
0.4
0.2
0
200
400
600
800
1000
1200
1400
1600
MAP POSITION
chloramphenicol^{r} gene

15d


Temperature 77.8 °C
PROBABILITY
1.0
0.8
0.6
0.4
0.2
0
200
400
600
800
1000
1200
1400
1600
MAP POSITION
chloramphenicol^{r} gene

15e

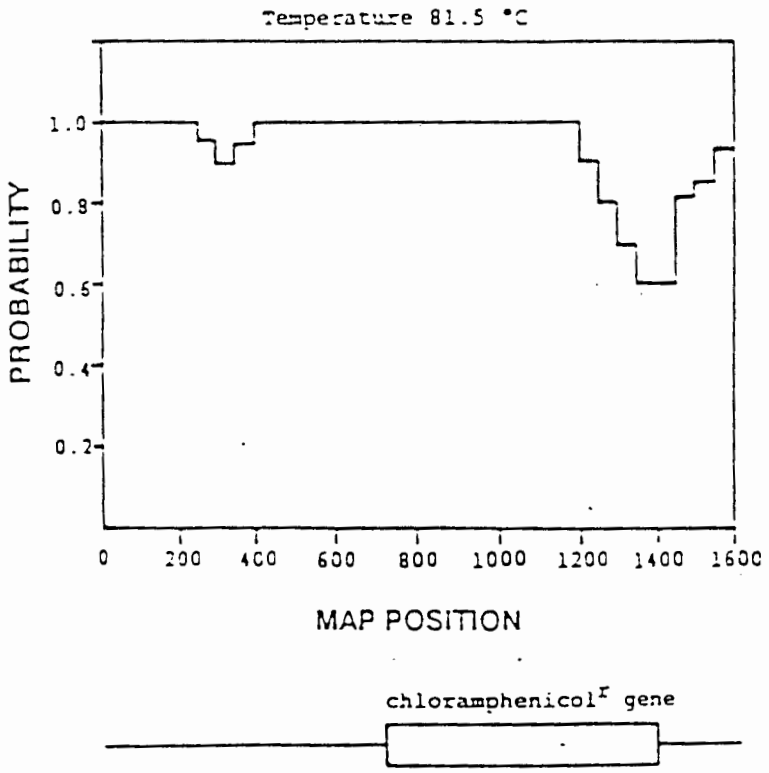


Fig 15. Melting maps of **pGV403** DNA obtained in 0.1*SSC from EM. a) 70.2°C b) 71.0°C c) 73.8°C d) 76.0°C e) 79.5°C. Probability represents the likelihood that a given region of the plasmid is denatured (single-stranded) at the specified temperature. Probability of 0 refers to double-stranded DNA, whereas at probability equal to 1.0, DNA is denatured. Temperatures represent the fixing temperatures.

VI.2 plasmid pUC9

The linearized form of **pUC9** (*EcoRI* digest) was fixed using the same method as described for **pGV403**. The melting curve of **pUC9** (Fig 16) shows 8 clearly visible transitions (peaks). At seven different temperatures (indicated by arrows) samples were taken and fixed by adding glyoxal.

The results obtained from **EM** (Fig 17) were digitized and denaturation maps were derived (Fig 18).

Fig 16

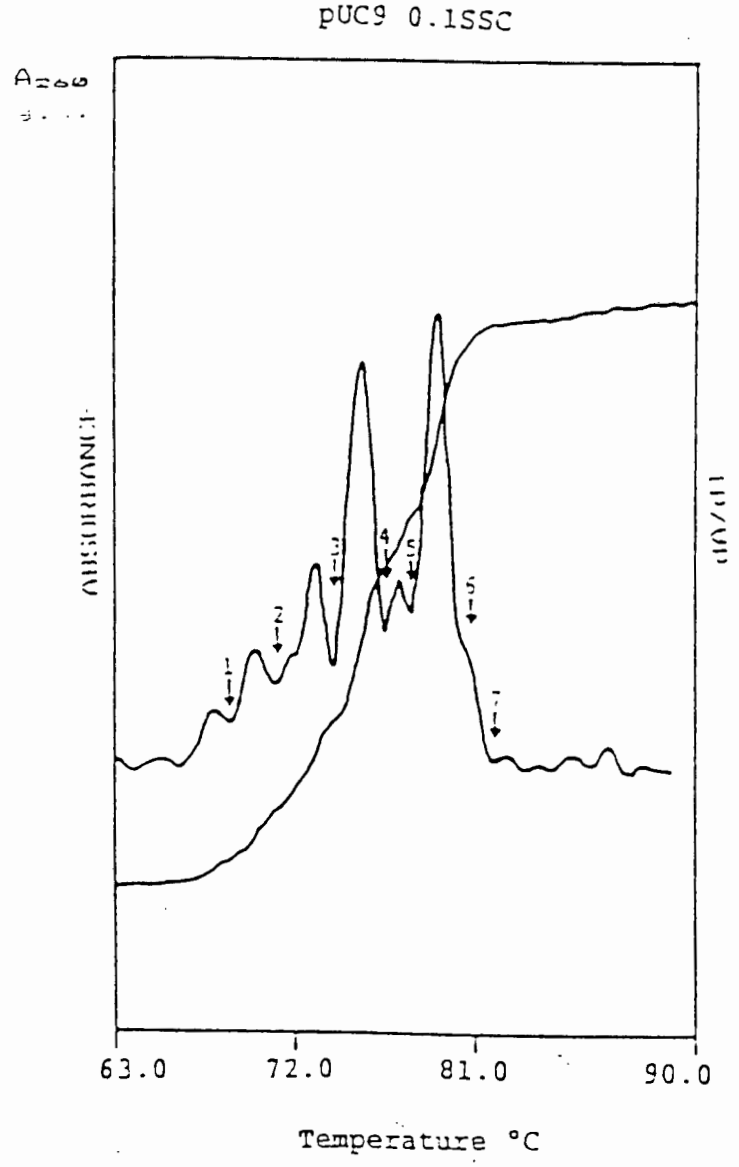


Fig 16. Thermal denaturation curve and first derivative for **pUC9** in 0.1*SSC. Arrows indicate the temperature of fixation of partially denatured DNA. 1) 70.2°C 2) 71.0°C 3) 73.8°C 4) 76°C 5) 77.3°C 6)79.5°C 7) 81.0°C

Fig 17

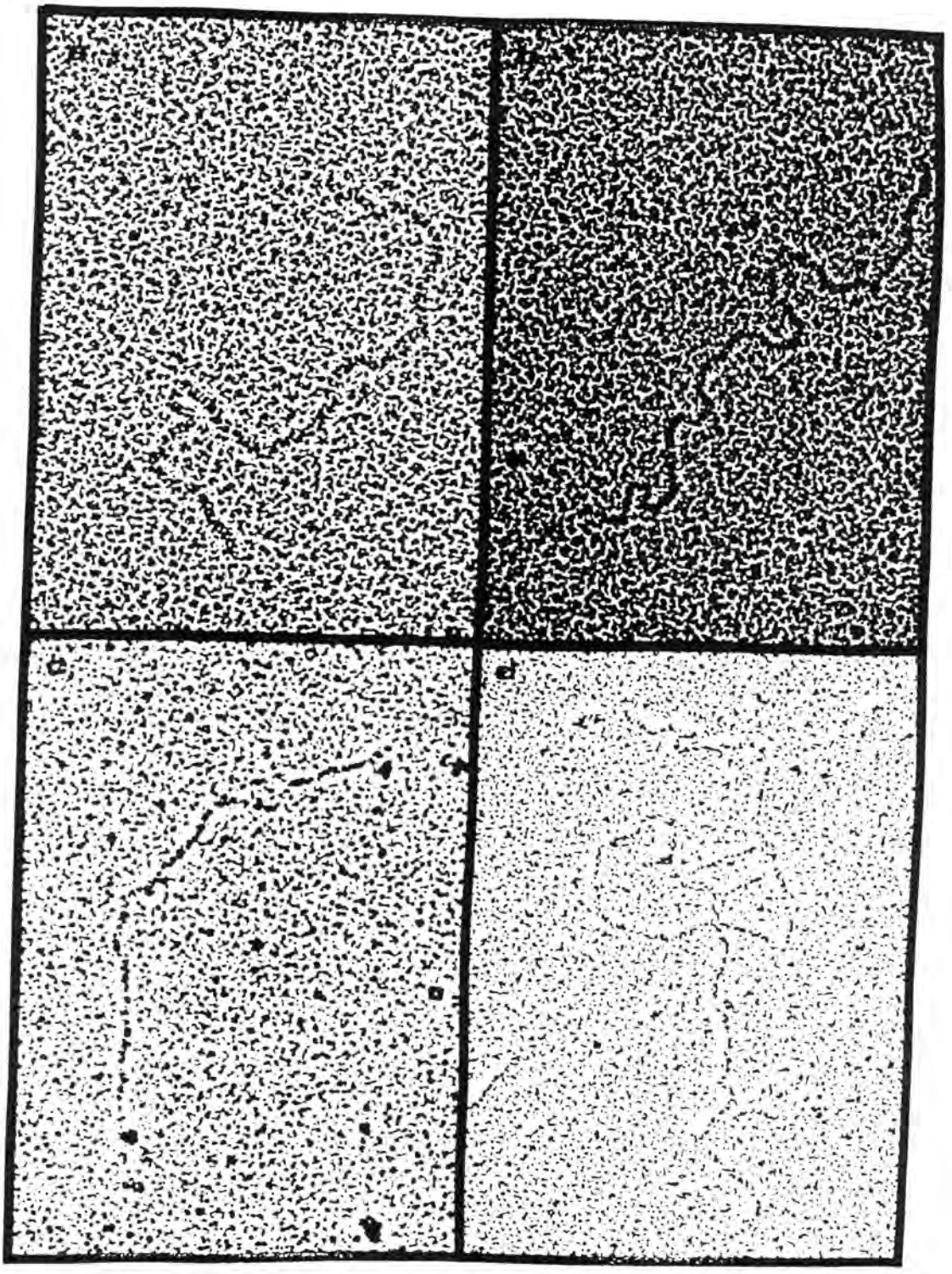
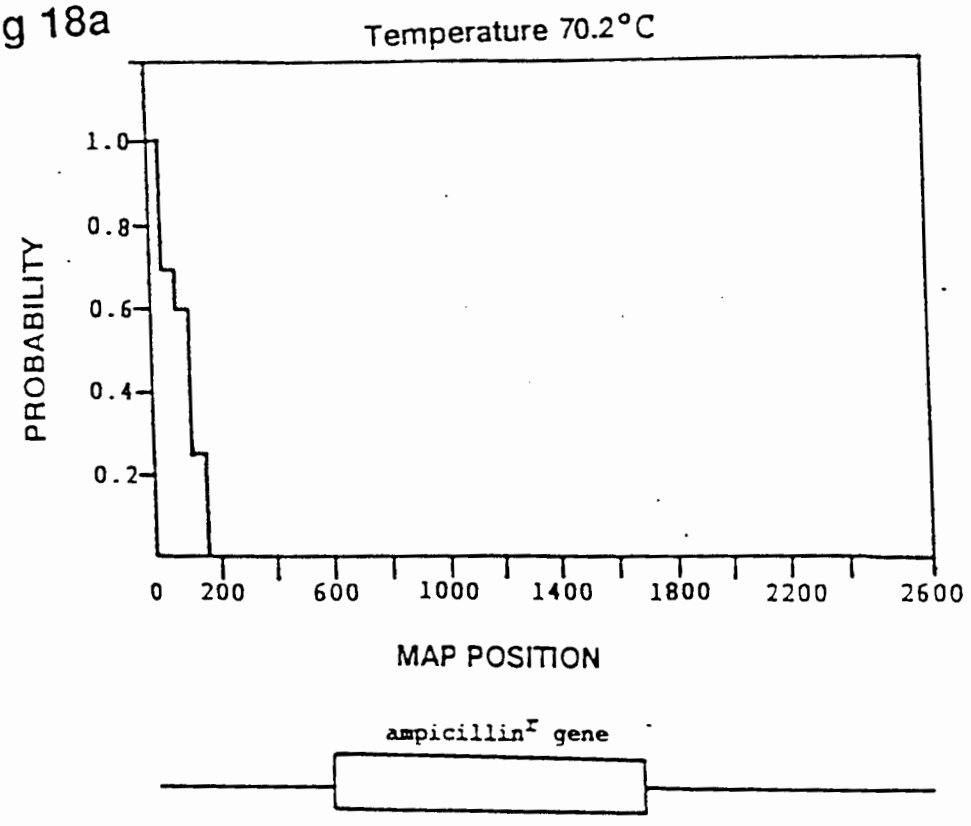
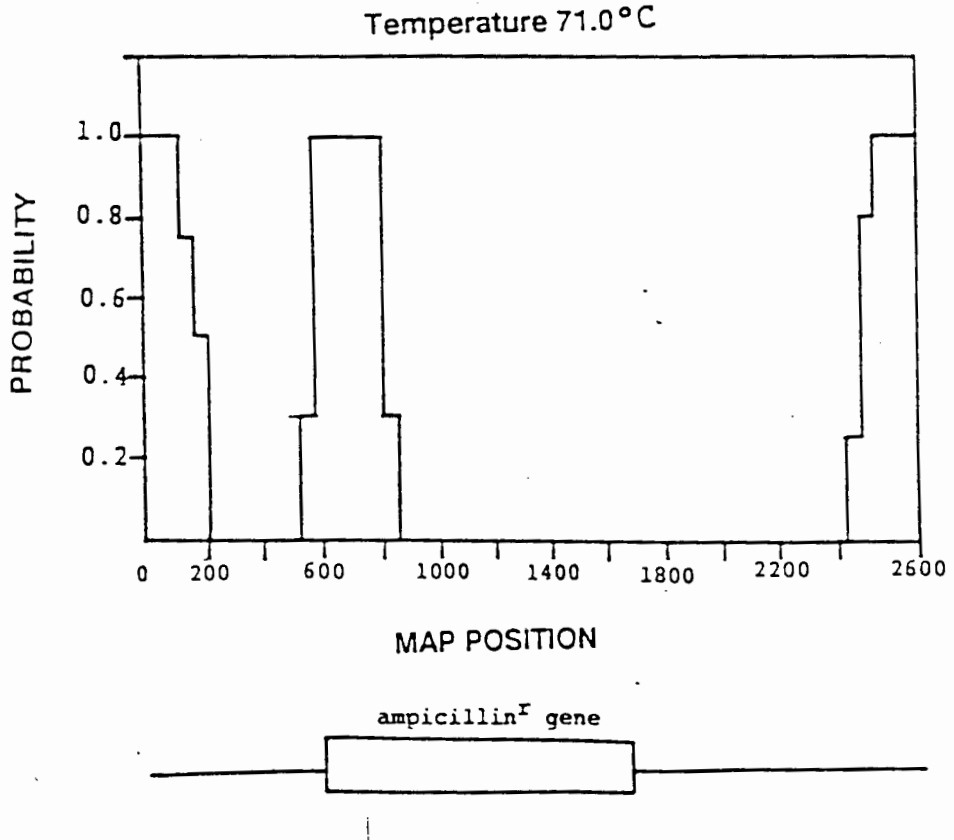


Fig 17. Electron micrographs of partially denatured **pUC9** plasmid DNA. a) 70.2°C b) 71.0°C c) 73.8°C d) 76.0°C Micrographs presents only well defined stages which were easy to observe on electron micrographs.

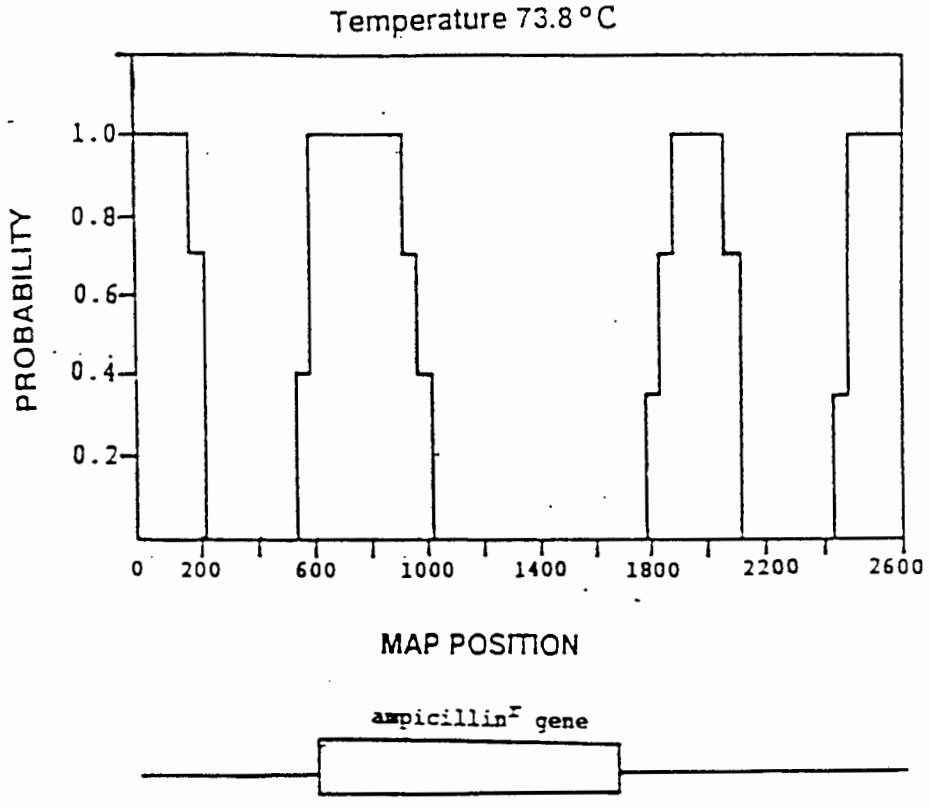
Fig 18a


Temperature 70.2°C
PROBABILITY
1.0
0.8
0.6
0.4
0.2
0
200
600
1000
1400
1800
2200
2600
MAP POSITION
ampicillin^r gene

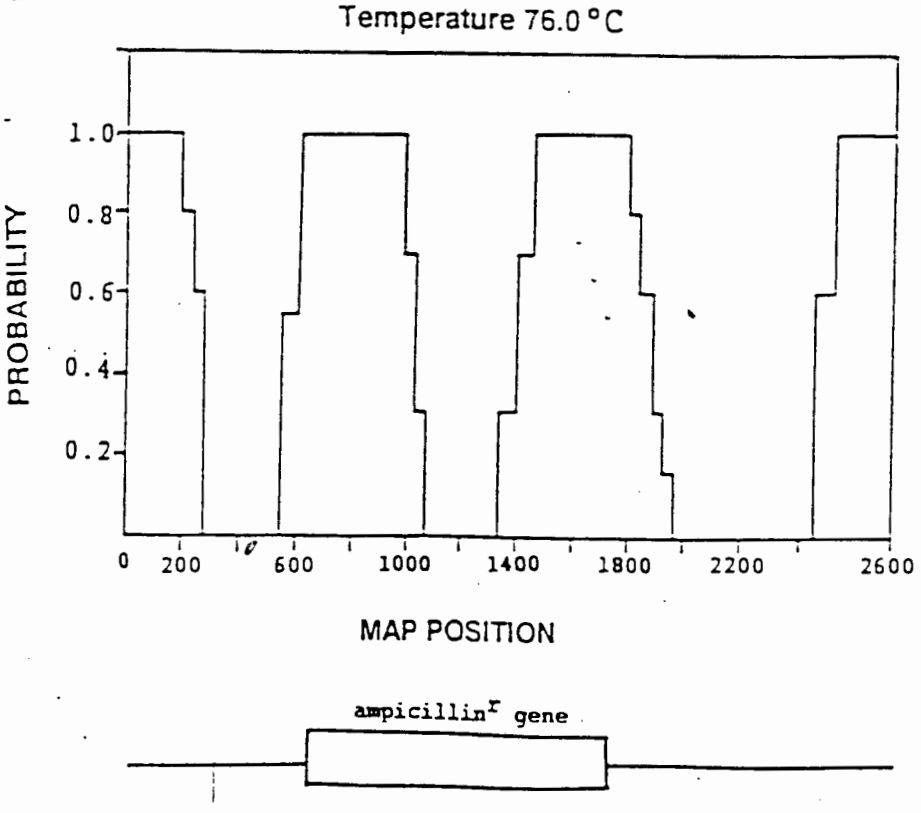
18b


Temperature 71.0°C
PROBABILITY
1.0
0.8
0.6
0.4
0.2
0
200
600
1000
1400
1800
2200
2600
MAP POSITION
ampicillin^r gene

18c


Temperature 73.8 °C
PROBABILITY
1.0
0.8
0.6
0.4
0.2
0
200
600
1000
1400
1800
2200
2600
MAP POSITION
ampicillin^r gene

18d


Temperature 76.0 °C
PROBABILITY
1.0
0.8
0.6
0.4
0.2
0
200
600
1000
1400
1800
2200
2600
MAP POSITION
ampicillin^r gene

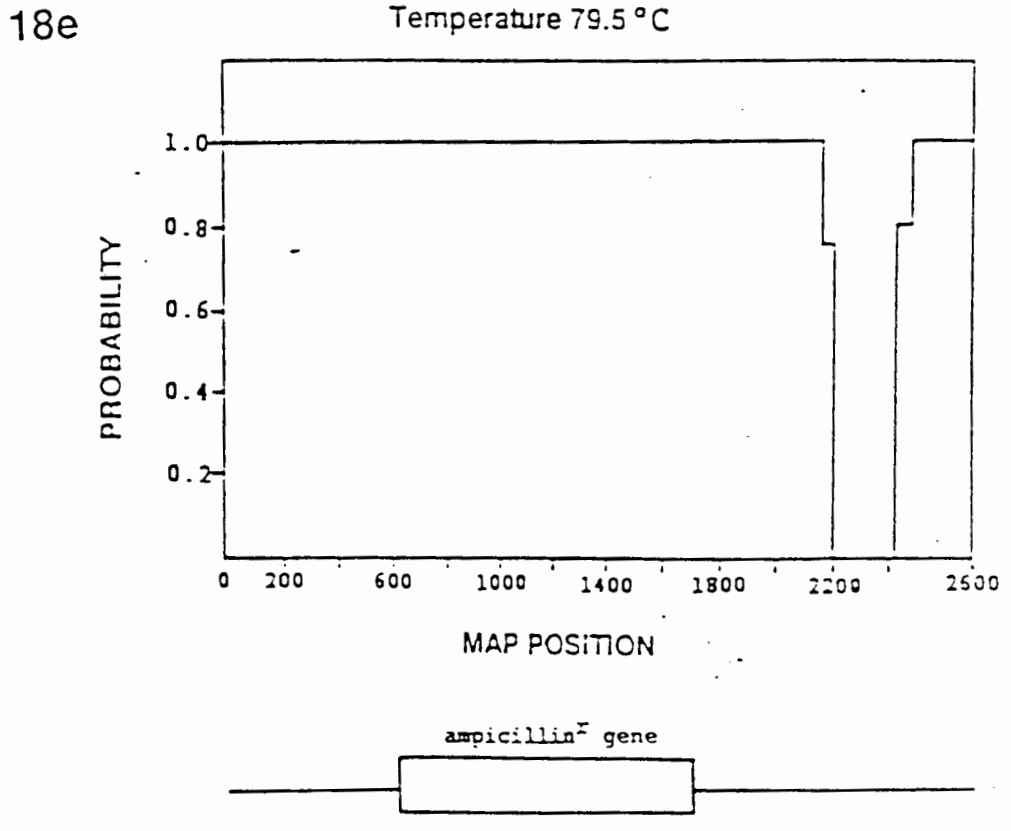


Fig 18. Melting maps for **pUC9** DNA obtained in 0.1*SSC from EM a) 70.8°C b) 72.2°C c) 76.4°C d) 77.8°C e) 81.5°C. Probability represents the likelihood that a given region of the plasmid is denatured (single-stranded) at the specified temperature. Probability of 0 refers to double-stranded DNA, whereas at probability equal to 1.0, DNA is denatured. Temperature represents the fixing temperature.

Summary and conclusion of the analysis of intermediate states of DNA analyzed by E.M.

The results presented in this chapter are in good agreement with those previously published for **pBR322**, e.g. by Lyubchenko *et. al.*[16,21,39]. Melting of DNA clearly occurs in a step wise fashion suggesting that in the DNA molecule there exist regions with definitive thermal stability (cooperative melting regions). Regions with lower GC content melt at lower temperatures and regions with higher GC content melt at higher temperatures. Each peak of the differential melting curve corresponds to the melting of one or possibly more that one DNA sequence. Analysis of E.M. denaturation maps shows that DNA regions melt cooperatively. The cooperativity in the melting curves is defined very well by the pointed shape of the peaks (dA/dT).

The EM results clearly show that in both plasmid families melting begins at free ends since there is lower stacking energy (note: arrows, Fig 14b, indicating the forks at the one or both ends of plasmid). It can also be observed that the single stranded regions (loops) created during the melting process were generally quite well isolated from one another. As was mentioned in the introduction the melting experiments can effectively simulate the biological processes which involve DNA, (e.g. the binding and action of RNA polymerase [26]). These processes can be visualized by electron microscopy of partly melted DNA. This technique is also able to show that melting is initiated at specific regions and spreads in discrete steps until the entire molecule is denatured. This offers a

partial explanation for the origin of "peaks" recorded by DSC or by UV-spectroscopic thermal denaturation scans.

CHAPTER VII

Thermal Denaturation Studies of Genes.

Introduction

After investigating the average features of unfractionated plasmid sequences the same approach was applied to well-defined regions (genes) contained within the plasmids studied earlier to ascertain whether it is possible to assign a portion of the complex DSC melting profile or thermal denaturation as monitored by UV spectroscopy of the whole plasmid to a given gene.
A gene can be regarded as a basic biological unit of genomic DNA and as such may have its own unique physical properties [51]. This can also be concluded from the results shown in previous chapters.

The most fundamental biological processes require DNA to be temporarily in a single stranded form (e.g. transcription, replication). It is of a considerable advantage that regions which are required to form single strand intermediates should not be too stable from a thermodynamic point of view. For some time it has been known that ***RNA*** polymerase causes local strand separation which takes the form of small internal "loops" similar to those obtained during melting experiments. This is another indication for the importance of thermodynamic properties as very important factors during such biological processes. This falls in line with the observation that the promoter region of genes

rich which leads to the conclusion that relative structural stability plays an important role during the transcription process.

Purified gene sequences were isolated from three plasmids, each with unique genetic properties, namely **pUC9**, **pGV403**, and **pBR322**. The thermal denaturation of these purified gene sequences, the Ampicillin resistance gene, the Chloramphenicol resistance gene, and the Tetracycline resistance gene, will be demonstrated in the following chapter.

VII.1. Isolation and purification of selected gene sequences

All genes investigated were purified on an HPLC column NUCLEOGEN 400-7DEAE with elution buffers A: 20mM Phosphate buffer, 6M urea, pH 6.7, buffer B: Buffer A + 1.2M KCl. The gradient applied varied from 840mM to 1200mM KCl over 90 min at a flow rate of 2 ml/min.

VII.1.1 Purification of a Chloramphenicol resistance gene from plasmid pGV403

The Chloramphenicol resistance gene was isolated from **pGV403** restriction endonuclease *RsaI* digest. Fig 19 shows the gel pattern of the digest.

Restriction fragments were purified on the HPLC column as described above. Fig 20 shows the elution profile of the efluent from the column. Electrophoresis of the fractions showed that the gene sequence eluted after approximately 60 min. The fraction was collected, dialyzed overnight and precipitated by adding 100% ethanol (3 times the initial volume).

The purity of the gene was checked on the 1% agarose gel. (Fig 19, lane 3).

Fig 19

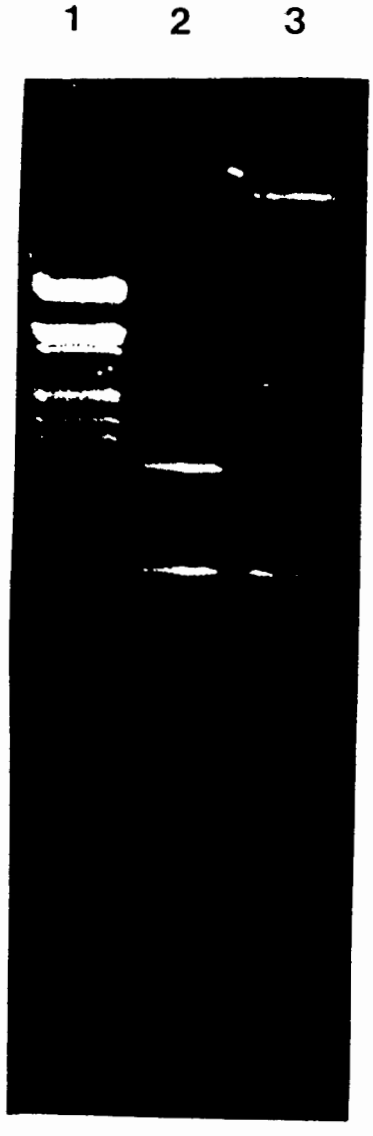


Fig 19. Restriction enzyme digest of **pGV403**
1% agarose gel electrophoresis. Lane 1 molecular weight marker λDNA (***EcoRI***, ***HindIII*** digest). Lane 2 ***RsaI*** digest of **pGV403** plasmid. Lane 3 restriction fragment containing Chloramphenicol resistance gene.

Fig 20

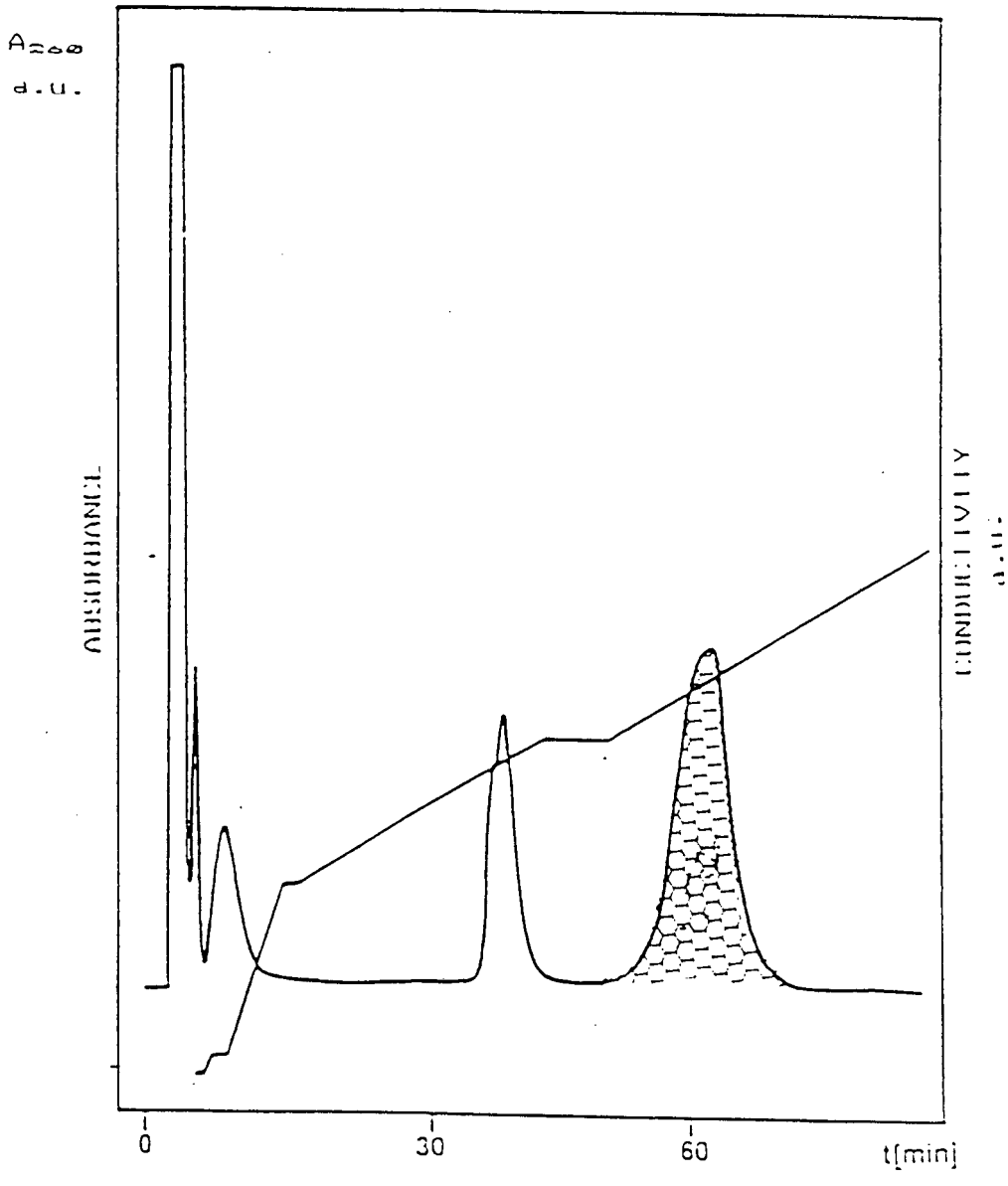


Fig 20. Elution profile of the Choramphenicol resistance gene.
Column: NUCLEOGEN 400-7DEAE
Sample: ***RsaI*** digest of **pGV403** plasmid
Eluant: KCl gradient in 20mM Phosphate buffer, 6M urea.
Flow rate: 2 ml/min

VII.1.2. Purification of a Ampicillin resistance gene from plasmid pUC9

The Ampicillin resistance gene was separated from **pUC9** vector by using the restriction endonucleases *SspI*, *PvuII*, and *BglI*.

The result of the digestions was checked on an agarose gel (Fig 21).

Fig 21

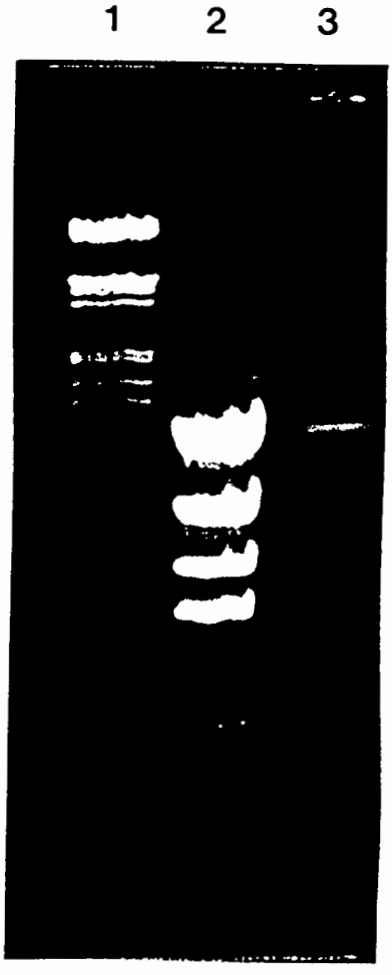


Fig 21. Restriction enzyme digest of **pUC9** plasmid.
1% agarose gel electrophoresis. Lane 1 molecular weight marker λDNA (***EcoRI***, ***HindIII***, digest). Lane 2 ***SspI***, ***PvuII***, and ***BglI*** digest. Lane 3 restriction fragment containing the Ampicillin resistance gene.

The fragment containing the gene was purified on a HPLC column as described before. The gene-containing fraction eluted after approximately 60 min at 68% of buffer B.

Fig 22 shows the elution pattern of the Ampicillin resistance gene from **pUC9**.

Fig 22

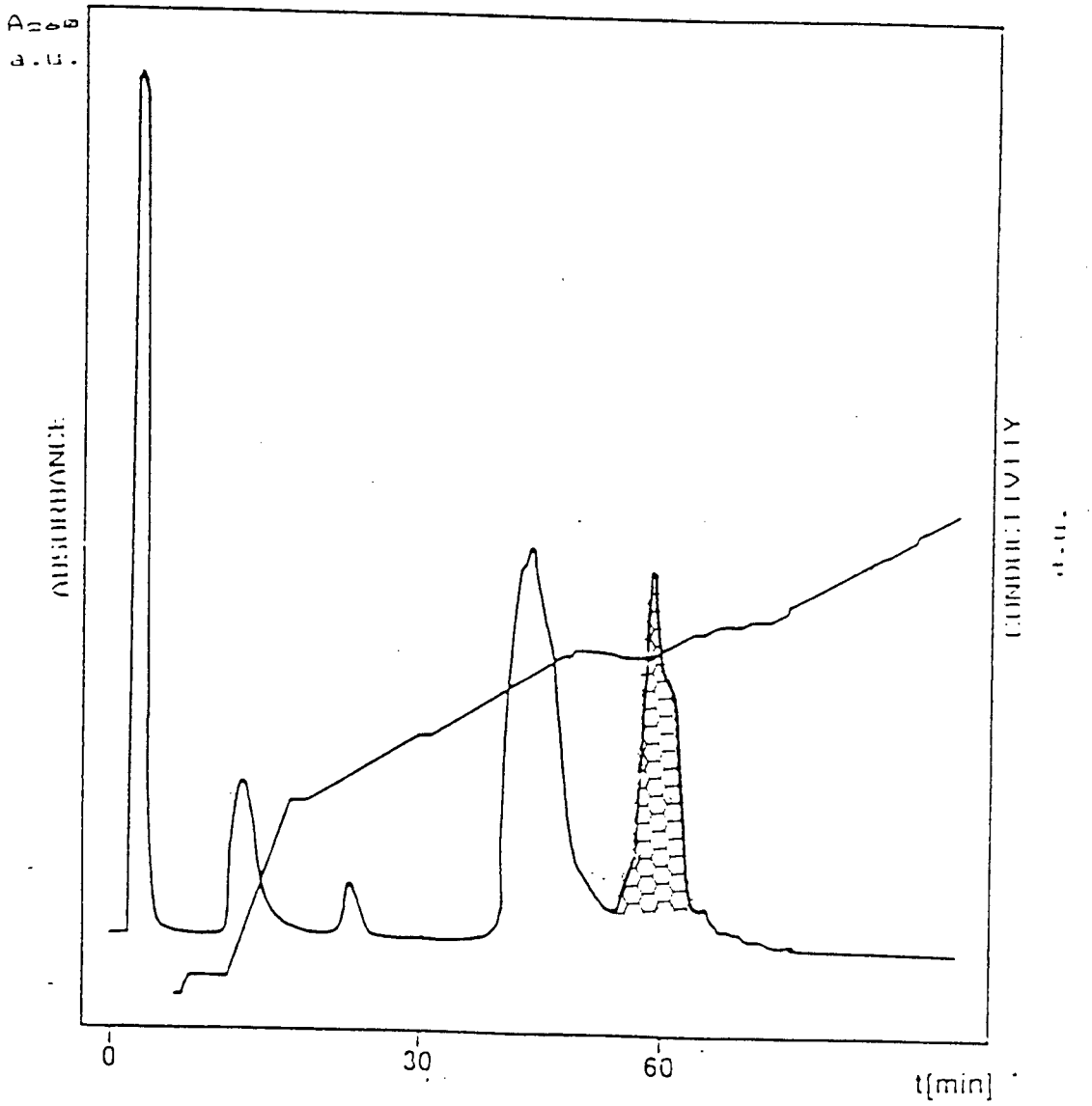


Fig 22. Elution profile of the Ampicillin resistance gene.
Column: NUCLEOGEN 400-7DEAE
Sample: ***SspI PvuII BglI*** digest of **pUC9** plasmid
Eluant: KCl gradient in 20mM Phosphate buffer, 6M urea.
Flow rate: 2 ml/min

VII.1.3 Purification of a Tetracycline resistance gene from plasmid pBR322

The Tetracycline resistance gene was released from **pBR322** vector by *EcoRV* and *AvaI* endonuclease digestion. The result of the digest was checked on an agarose gel (Fig 23).

Fig 23

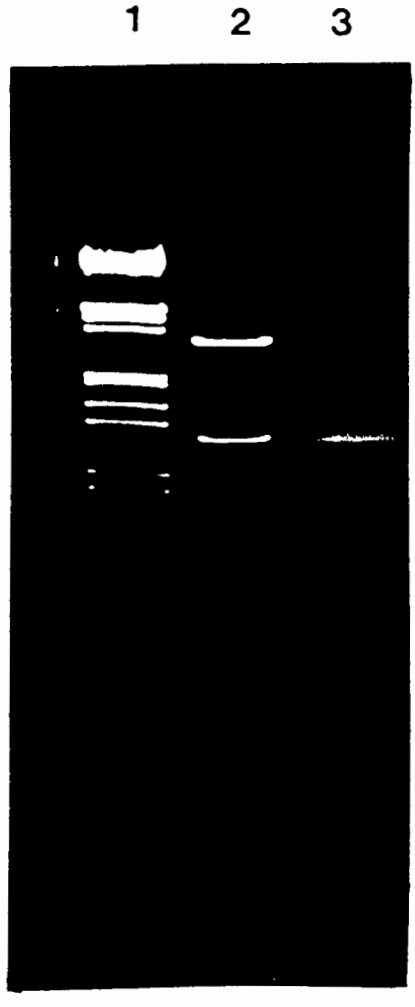


Fig 23. Restriction enzyme digest of **pBR322** plasmid.
1% agarose gel electrophoresis. Lane 1 molecular weight marker λDNA (*EcoRI*, *HindIII* digest). Lane 2 *EcoRV*, *AvaI* digestion. Lane 3 restriction fragment containing the Tetracycline resistance gene.

The restriction fragments were purified as previously described on an HPLC column and the fraction containing the gene was collected at approximately 60 minutes (Fig 24).

Fig 24

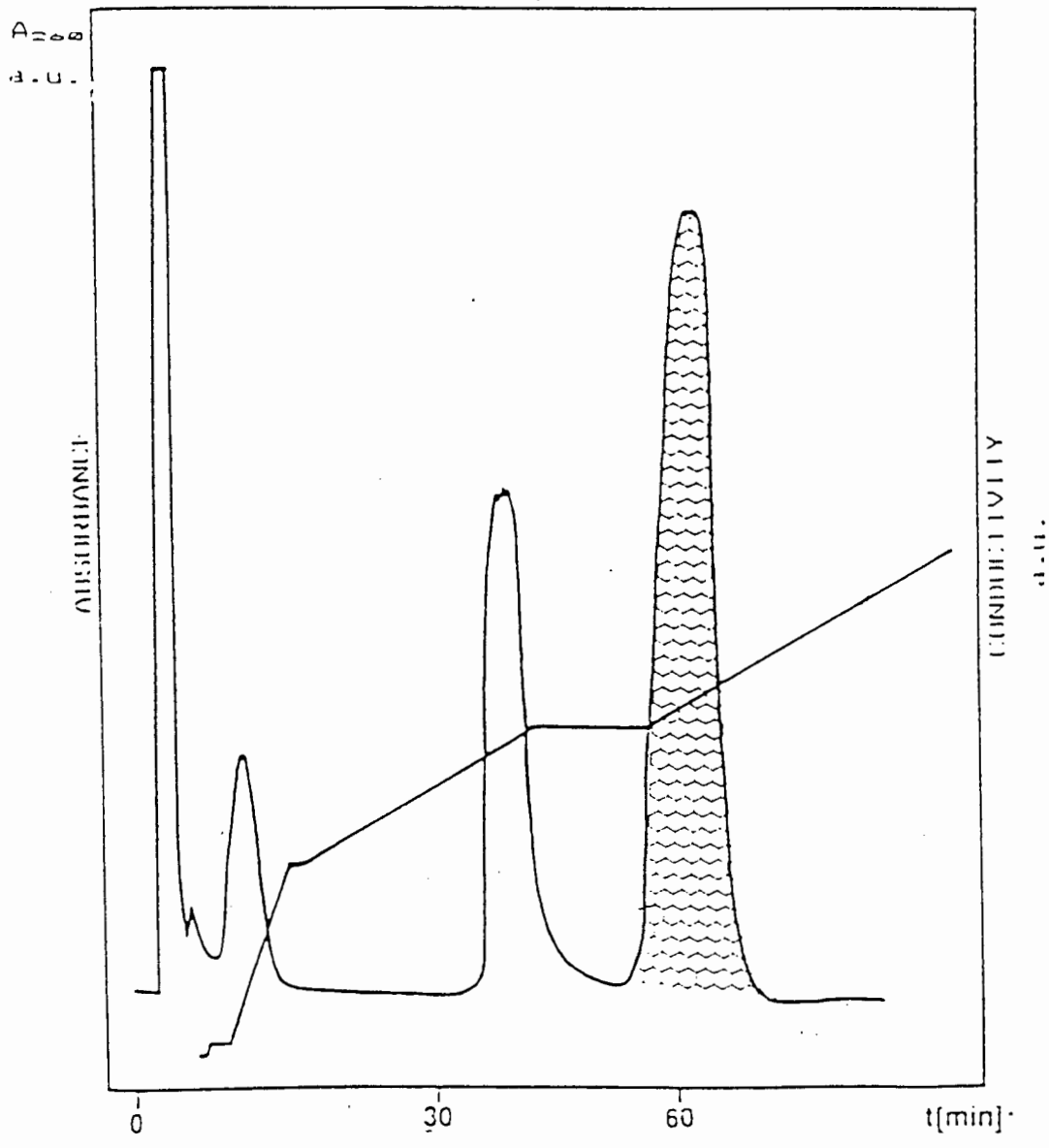


Fig 24. Elution profile of the Tetracycline resistance gene.
Column: NUCLEOGEN 400-7DEAE
Sample: *EcoRV*, *AvaI* digest of **pBR322** plasmid
Eluant: KCl gradient in 20mM Phosphate buffer, 6M urea.
Flow rate: 2 ml/min

VII.1.4. Thermal denaturation of selected genes

All sequences containing the particular genes were thermally denatured over the temperature range from 38°C to 98°C at a heating rate of 1°C per minute and the absorbance change (hyperchromic shift) was monitored at 260nm.
As a control unfractionated linearized plasmid DNA was denatured under the same set of experimental conditions.

VII.1.4.a. Thermal denaturation of a Chloramphenicol resistance gene from plasmid pGV403.

The Cholamphenicol resistance gene was denatured according to the methods described above and compared with the melting profile of the parental vector. The relevant results (ΔH_{VH} versus temperature t_m) are listed in Table 11. Data were derived from Fig 25 a, b.

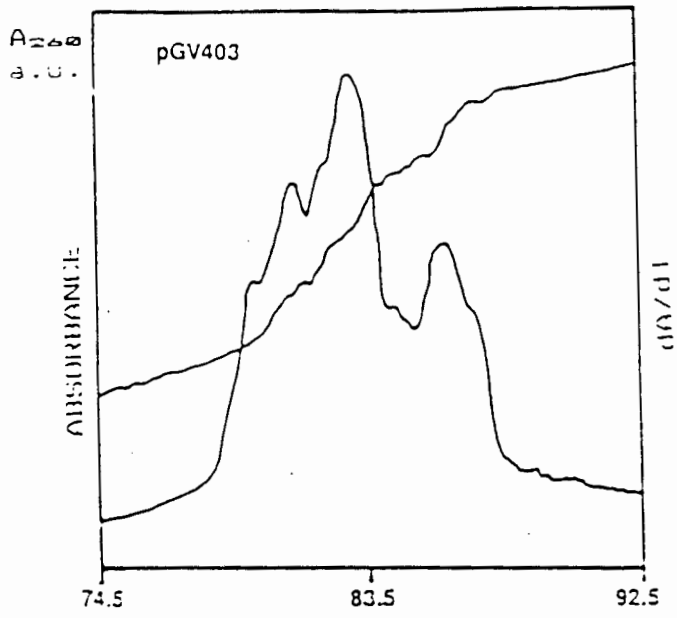
Table 11

Conformational transitions and ΔH_{VH} of **pGV403** and its Chloramphenicol resistance gene sequence

	t_m [°C]	ΔH_{VH} of plasmid [kJ/mol]*10^3	t_m [°C]	ΔH_{VH} of gene [kJ/mol]*10^3
1	77.5	2.20	77.5	2.17
2	79.1	3.10		
3	80.4	4.22		
4	82.0	5.80	82.0	5.70
5	84.6	1.70		
6	86.6	2.91		
7	88.3	1.80		

It is evident that ΔH_{VH} calculated for the transitions 1 and 4 is in good agreement with corresponding results for the gene and for the parental plasmid. This implies that some regions exist which have conserved thermodynamic properties within the plasmid and that these regions coincide with the gene sequences studied.

Fig 25a



25b

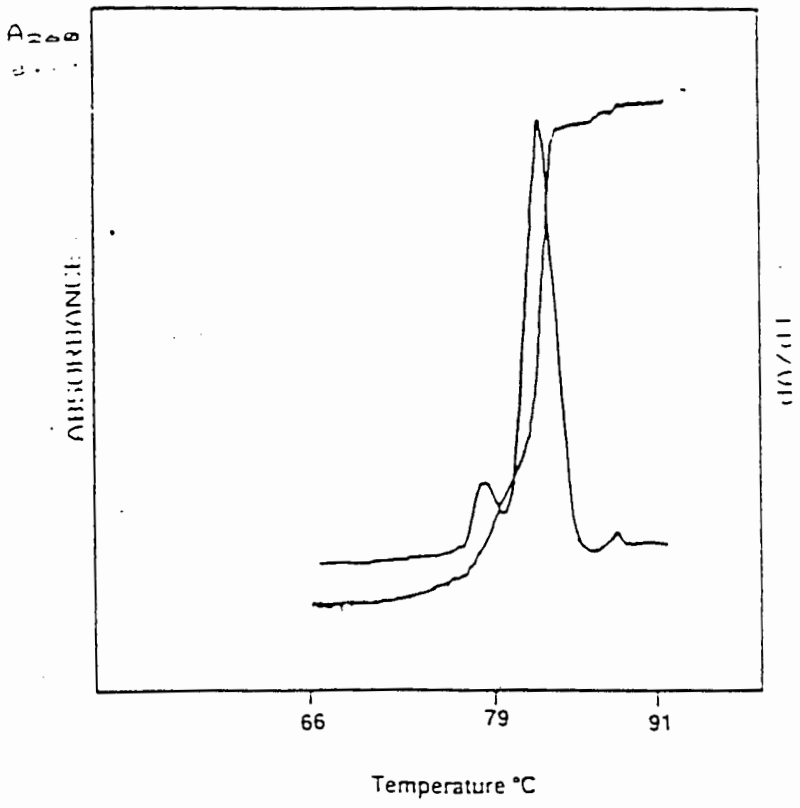


Fig 25. Thermal denaturation profile and first derivative plot a) **pGV403** plasmid b) Chloramphenicol resistance gene
Heating rate 1°C per minute. Buffer 50 mM NaCl, 0.5 mM EDTA, 1 mM TRIS pH 7.6

VII.1.4.b. Thermal denaturation of an Ampicillin resistance gene from plasmid pUC9

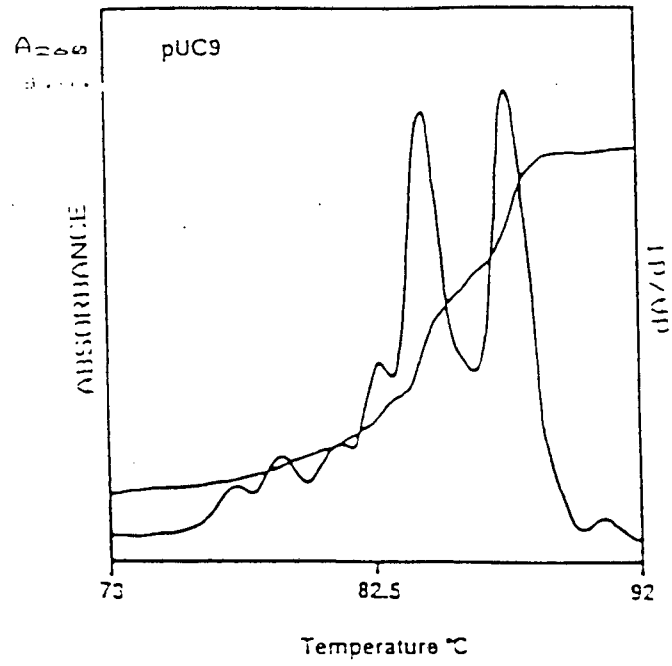
The Ampicillin resistance gene was isolated and denatured according to the methods described above and compared with the melting profile of the parental vector. The results (ΔH_{VH}, temperature t_m) are listed in Table 12. The data were derived from Fig 26 a, b.

Table 12

Conformational transitions and ΔH_{VH} of **pUC9** and its Ampicillin resistance gene sequence

t_m [°C]	ΔH_{VH} of plasmid [kJ/mol]*10^3	t_m [°C]	ΔH_{VH} of gene [kJ/mol]*10^3
76.6	2.02		
78.4	4.95	78.6	3.93
80.8	1.52		
82.4	3.91	82.4	3.68
83.8	7.92		
85.8	5.34	85.8	5.22
87.4	4.62		
89.0	8.15	89.3	3.20

Fig 26a



26b

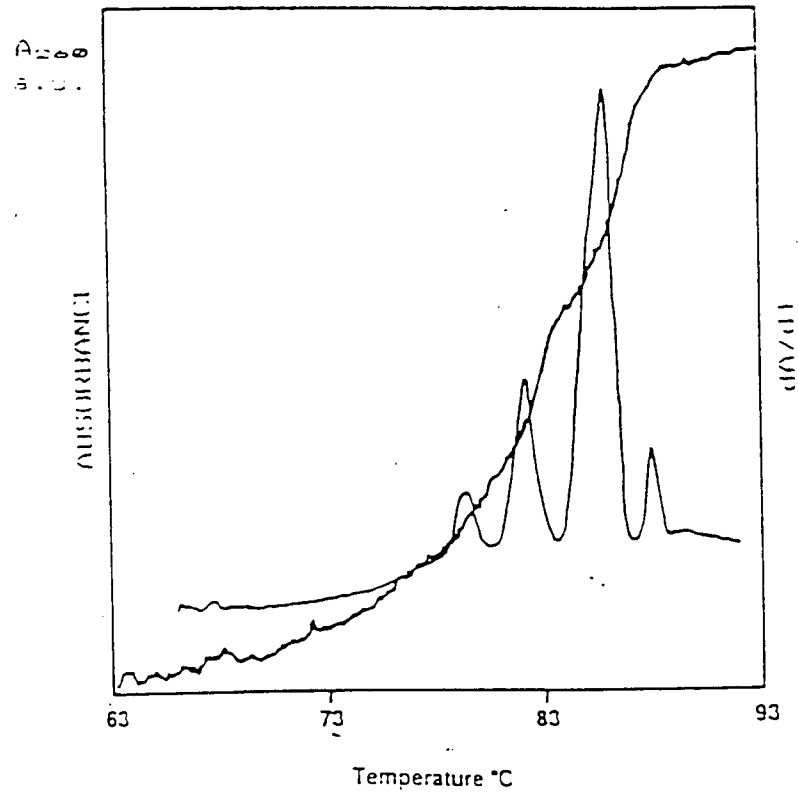


Fig 26.Thermal denaturation profiles and first derivative plot a) **pUC9** plasmid b) Ampicillin resistance gene
Heating rate 1°C per minute. Buffer 50 mM NaCl, 0.5 mM EDTA, 1 mM TRIS pH 7.6

VII.1.4.c. Thermal denaturation of a Tetracycline resistance gene from plasmid pBR322

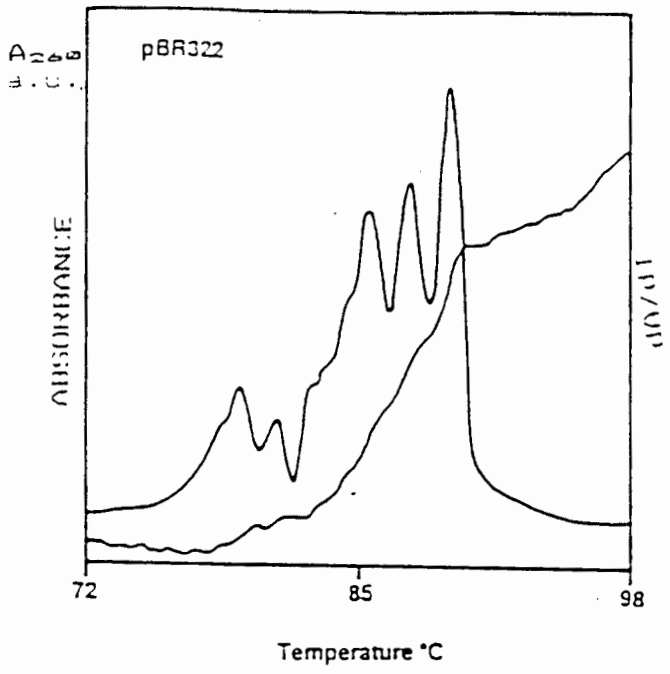
The Tetracycline resistance gene was isolated and denatured according to the methods described above and compared to the melting profile of the parental vector. The results (ΔH_{VH}, temperature t_m) are listed in Table 13. The data were derived from Fig 27 a, b.

Table 13

Conformational transitions and ΔH_{VH} of **pBR322** and its Tetracycline resistance gene sequence

t_m [°C]	ΔH_{VH} of plasmid [kJ/mol]*10^3	t_m [°C]	ΔH_{VH} of gene [kJ/mol]*10^3
73.4	3.70		
77.0	4.26		
78.0	5.00		
80.4	2.53		
82.2	5.12		
84.4	2.22	84.5	2.78
85.8	5.27		
87.2	3.16	87.2	3.55
88.2	3.97		

Fig 27a



27b

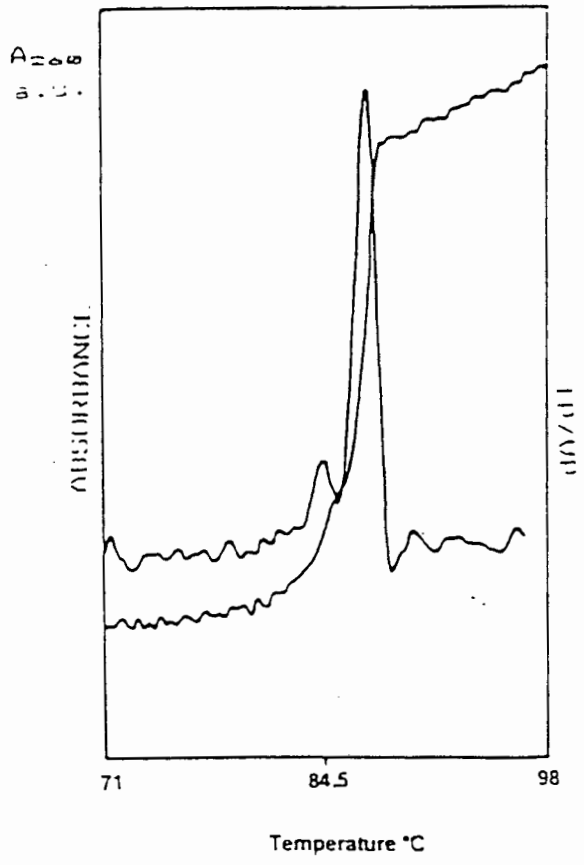


Fig 27. Thermal denaturation profiles and first derivative plot a) **pBR322** plasmid b) Tetracycline resistance gene
Heating rate 1°C per minute. Buffer 50 mM NaCl, 0.5 mM EDTA, 1 mM TRIS pH 7.6

Summary and conclusion of the thermal denaturation studies of selected genes

---

The main aim of this study was to compare the thermal denaturation profile of important biological units, i.e. genes, with the denaturation profile of the parental vectors and to establish which transitions ("peaks") of the complex multiple transition occurring within the plasmid can be assigned to the gene sequence.

Studies were done on three different plasmids: **pGV403**, **pUC9** and **pBR322** which contain the genes for Chloramphenicol resistance, Ampicillin resistance and Tetracycline resistance respectively. The results of the experiments have been summarized in **Tables 11** for **pGV403** and the Chloramphenicol resistance gene, **12** for **pUC9** and the Ampicillin resistance gene, and **13** for **pBR322** and the Tetracycline resistance gene. It can be seen that the Chloramphenicol resistance gene melts in two steps at t_ms 77.5°C and 82.0°C. This corresponds to the first and fourth transition peak of the **pGV403** melting profile. The Ampicillin resistance gene was seen to melt in four steps, at t_ms of 78.6°C, 82.4°C, 85.8°C and 89.3°C, which correspond to the second, fourth, sixth and eighth step of the melting curve of the parental plasmid **pUC9**. The Tetracycline resistance gene melts in two steps at t_ms 84.5°C and 87.2°C which corresponded to the sixth and eighth peak of the melting profile of plasmid **pBR322**.

There are observable differences in the melting temperatures taken from the electron micrographs and from the DSC or UV-spectrophotometric results listed in this chapter due to different solution conditions.

The results given in this chapter support previous findings that genes reveal a quite distinct melting patterns (which correlates well with EM micrographs) with well defined melting temperatures and transition enthalpies ΔH_{VH}. This also holds for the parental plasmid (regarded here as a "mini-genome"). Therefore, it is reasonable to conclude that genes can be characterized by their melting temperature, melting profile and the melting enthalpy (ΔH_{VH}).

CHAPTER VIII

Conclusions

This study was set up to answer the two questions posed in the introduction, viz.

1. Is there a correlation between the base sequence function and a measurable physical property which can be assigned to biologically important units such as promoter or coding sequences?

2. Is there a correlation between the denaturation of gene sequences and cooperative transitions observed in a given temperature interval?

The results lead to tentative conclusion that the answer to both of the questions should be "yes". How can one be certain of this?

In order to answer these questions three different experimental techniques were applied, namely differential scanning calorimetry, UV-spectrophotometry and electron microscopy. The reason for using three different experimental techniques was that calorimetry measures the total enthalpy of disrupting base-pairing and base stacking, while the scanning of the change of absorbance (UV-spectrophotometry) reflects mainly changes in base stacking interactions. Electron microscopy is the

only method which allows one to arrest intermediate states and to visualize the physical changes occuring during elevating the temperature. A combination of these methods certainly yields more information on the physical properties of DNA than each alone.

The results presented in chapter II derived from differential scanning calorimetry show that within the two families of plasmids studied viz. (1) **pGV403** and its derivatives **pHP2** and **pHP32**, and (2) **pUC9** and its derivatives **pBS** and **pBR322** there are highly conserved features in the denaturation profiles. As listed in Tables 5 and 6, the corresponding "peaks" (transitions) have the same t_ms and transition enthalpies (ΔH) within the family, e.g. for the **pGV403** family, peak number I is common to all three plasmids whereas peak number III of **pGV403** corresponds to peak number IV of **pHP2** and of **pHP32** plasmids. In the **pUC9** family peak number II is common to all members of this family. Peak number IV of **pUC9** corresponds to peak number IV of **pBS** and peak number V of **pBR322** plasmid. Finally peak number VI is common for **pUC9** and **pBS** sequence and corresponds to peak number IX of **pBR322** DNA.

As mentioned above differential scanning calorimetry (DSC) measures the total enthalpy of disruption of base pairing and base stacking. The results of these measurements and the calculated thermodynamic state functions suggest that within both groups of plasmids some sequences exist with very conserved thermodynamic properties which are independent of the neighbouring sequences. In other words there are cooperative melting regions, defined by their endothermic peaks

(transitions) which correspond most probably to defined biological properties.

In order to verify the results obtained by differential scanning calorimetry (DSC), high resolution thermal denaturation profiles of both plasmid families were studied with the help of UV-spectroscopy.

There are obvious reasons for avoiding DSC as the only technique. The DASM-4 calorimeter (NPO "Biopribor" Academy of Science USSR)[47] is most accurate for quantities of DNA corresponding to 6 to 9 A_{260} units (300-450 g) of DNA. It has been shown that melting patterns are slightly dependent on the concentration of the DNA and sometimes it is simply very difficult to purify sufficient material. The complementary physical technique to cope with the lower concentration is UV-spectroscopy.

As was expected the thermal denaturation profiles obtained by this technique were similar but not identical to DSC profiles because of the nature of the physical signal.

As was found for the melting profiles obtained from spectroscopic measurements both plasmid families reveal some conserved features which are listed in Tables 9 and 10. For the **pGV403** family, peaks I and IV are common to all profiles and for the **pUC9** family peaks II and IV are redundant. Peak VI of the denaturation profile of **pUC9** corresponds to the peak number VI of **pBS** DNA and to the peak VII of **pBR322** DNA respectively.

A slight discrepancy is observed for the plasmid melting temperatures derived from differential scanning calorimetry as compared to high

resolution thermal denaturation. This is most probably due to differences in the temperature registration.

It was mentioned in the **Introduction** that the average size of the cooperative melting unit is of the order of a hundred base pairs and the average length of exons, the "cooperative" genetic unit is of the same order. Thus it can be assumed that biologically important units, for instance prokaryotic genes, whose average length is of the order of a thousand basepairs may contain one or more cooperative melting units. It has been shown that a single gene can give rise to one or more transitions ("peaks") in a melting experiment, where each transition corresponds to a (physical) cooperative melting unit [14,19,46,53].

Based on this observation, experiments were designed to test this hypothesis and the results are presented in chapter VII. These studies were carried out using only one technique (UV-melting) for practical reasons (preparation of 500µg of "pure genes" was unjustified and unnecessary).

The results presented in Tables 11-13 support the basic hypothesis that a number of the transitions within the whole melting profile represent the cooperative melting unit present within the Ampicillin resistance gene, the Tetracycline resistance gene, and the Chloramphenicol resistance gene respectively.

In order to visualize the physical changes which occur at elevated temperatures samples were fixed with glyoxal and electron micrographs

of the partially denatured DNA were taken. The results are presented in Figs 15 and 18. They show that during the melting process internal loops in the DNA are created, which correspond to the comparative melting units. Their position relative to the restriction enzyme cutting site reflects the distribution of the local stability which is in agreement with the idea that the transient opening of the DNA double helix is required for biological function.

Blake *et.al.* [49,53] pioneering high resolution differential UV-melting procedures has established, *inter alia*, that there is a linear relationship between the GC content within the melting unit and the transition temperature of this unit.

Unfortunately only a few plasmids were studied and the main conclusion was derived mainly from plasmid **pBR322**. Chapter IV of this thesis is an attempt to verify these results for a larger family of plasmids, i.e. **pGV403** and its derivatives and **pUC9** and its derivatives. The results obtained here have confirmed those results obtained by Blake at.al. that melting units do melt according to their GC content.

In addition to the experimental observations a semi-empirical approach was developed to predict local sequence stability from nucleotide sequence data. The feature focused on in this approach is the local compositional complexity. The results obtained by calculating the local compositional complexity (LCC) within the chosen window (100 p) show that LCC in the gene area is rather smooth i.e. there are no abrupt

changes in the profile. This approach also allows one to calculate the thermodynamic properties of genes. LCC analysis shows that each change in the boundary sequences or changes within the base sequence itself caused by introduction of some "foreign" sequences of DNA into the gene are reflected in changes in the complexity pattern (Fig 10a-c). The analysis of the local compositional complexity gives a completely different view of the features of a given DNA sequence from the approach proposed by Volker Brendel [60] in "Towards the understanding the DNA as a language". His approach focuses on small units like base triplets, sextets, and up to octets and relates them to a language theory. LCC analysis treats DNA as a language composed of four different letters (basepairs ATGC) and calculates the local entropy within the sliding window along the DNA sequence. The results obtained by LCC confirm the results obtained in Chapter VII. Genes are unique structures with unique thermodynamic properties which can be defined by the local entropy (LCC).

It could well be that the correlations observed earlier [13,17,41,48] between melting profiles and the genetic maps are an accidental feature of the DNA. Therefore one of the tasks of this thesis was to test this possibility.

To this aim research was carried out on six different plasmids, which were divided into two groups. For each group a parental plasmid was chosen, resulting in a **pGV403** group and a **pUC9** group. Three different genes were present, namely the Chloramphenicol resistance gene for the plasmids of the **pGV403** family and the Ampicillin resistance gene and

the Tetracycline resistance gene for the plasmids of the **pUC9** family. Through comparison of the results obtained it is possible to conclude that there is a correlation between the physical and the genetic maps. Results presented here indicate that it is indeed possible to characterize genes by their thermodynamic properties (t_m, ΔH_{CAL}, ΔH_{VH}) and to show that they are most probably independent of the surrounding environment i.e. from surrounding sequences of nucleotides. The characteristic melting profile of a certain gene sequence could be used as an indicator for the presence of this particular gene in an unknown sequence.

In summary the following conclusion can be drawn: genes as biologically important units have a unique AT+GC content and arrangement, which will specify their biological as well as their physical properties. They are written in the ATGC language. This conclusion supports the hypothesis that evolution works inside the framework of thermodynamics and therefore genes exert their biological functions by mean of their physical properties.

CHAPTER IX

MATERIALS AND METHODS

IX.1. Preparation of plasmid DNA

Plasmids pUC9[43], **pGV403** [58], **pBR322** [30,31] and their derivatives were prepared in a three-step procedure (modified method of Clewel and Helinsky [9]):

Step I.

Ten milliliters of (in *E.coli* strain HB101) culture were prepared by inoculating 100µl of glycerol stock culture of the plasmid containing variant into 10ml L-broth (Maniatis *et. al.*[42]) with the required antibiotic present (50µg/ml Ampicillin for **pBR322**, **pUC9** and **pBS**, Chloramphenicol for **pGV403**, **pHP2** and **pHP32** respectively). The culture was incubated for 16 hours at 37°C with vigorous shaking. From this culture 2 ml was used to inoculate 1 liter of L-broth. This was incubated overnight (16 hours) at 37°C with vigorous shaking. The cells were harvested by centrifugation (15 minutes at 5,000 rpm, 4°C, GS3 Sorvall rotor).

Step II.

The pellets were separately resuspended in 8 ml of 50mM Tris-HCL buffer pH 7.6 and 1 ml of freshly made lysosyme was added from a 10 mg/ml stock solution. The mixture was incubated for 30 minutes at 4°C,

after which 1 ml of 500mM EDTA (pH 7.5) was added and incubation continued for 10 minutes. RNAse from a 10 mg/ml stock solution was added (0.1 ml) as well as 200µl of a 10% (v/v) Triton solution. The mixture was incubated at 4°C for 20 minutes after which centrifugation at 17000 rpm in a Sorvall SS34 rotor at 4°C was performed.
The supernatant was extracted once with phenol and twice with chloroform. The DNA from the aqueous phase was precipitated by addition of 2.5 volume of ice-cold ethanol and pellets were washed in 80% ethanol.
The DNA pellets were dried and resuspended in the following solution: 4.0 g CsCl, 4.2 ml TE, 200µl of EtBr (from the stock solution of 10mg/ml) per tube. The refractive index was adjusted to 13,920 by addition of extra water or CsCl as needed. The samples were centrifuged for 16 hours at 55000 rpm at 20°C in a VTi 65 rotor (Beckman).

Step III.
The supercoiled plasmid was collected by piercing the tube with a needle attached to a 1ml syringe (Maniatis [44]). EtBr was removed by repeated extraction with isoamyl alcohol and the solution was dialysed three times against 4 volumes of TE buffer. Plasmid DNA was precipitated by addition of 2.5 volumes of -20°C absolute ethanol. DNA was lyophilized and pellets were resuspended in TE buffer.

IX.2. Restriction endonuclease enzyme digestion

DNA was incubated with restriction enzymes at a ratio of 1 μg of DNA per 1 unit of enzyme with required buffer at 37°C for fifteen hours. The completeness of digestion was checked by agarose gel electrophoresis as described below. Restriction enzymes (EcoRI, PvuII, RsaI SspI, HindIII, BglI) and molecular weight marker were obtained from Boehringer-Mannheim.

IX.3. DNA Electrophoresis

Double-stranded fragments of DNA were analyzed on 1% agarose mini gel (according to Maniatis [44]) in TBE buffer. The agarose was suspended in TBE buffer, the mixture was then boiled, the gel poured into a mold and allowed to cool and set. The samples were dissolved in the following buffer: 50% v/v glycerol, 0.1 mM EDTA, 0.01% w/v bromophenol blue. Electrophoresis was performed at 100V constant voltage and monitored by ethidium bromide fluorescence on UV transilumination.

IX.4. HPLC (preparative) purification of DNA fragments

A NUCLEOGEN 4000-7 DEAE column was used for the isolation of DNA restriction fragments. Three hundred microliters (approximately 300μg of DNA) were taken and diluted with an equal volume of buffer (5M urea, 0.02M K-phosphate pH 6.5-7.0) and injected on to a column

equilibrated with a 40% solution of the following buffer **E**: 5M urea, 0.02M K-phosphate, 1.2M KCL pH 6.5-7.0. After injection an increasing gradient of buffer **E** was started (10mM/min). The restriction fragments were collected and dialysed against three changes of TE buffer 10 volumes for 1 hour each. The DNA restriction fragments were precipitated and analyzed on an agarose mini gel as described elsewhere (IX.1, IX.3).

IX.5. Electron Microscopy of partially melted DNA

IX.5.1. EM preparation and photography

Denatured samples of DNA were prepared according to Borovik *et. al.*[34]. Each denaturation was carried out as described below in a total volume of 1ml 0.1xSSC (0.015M NaCl, 0.0015 trisodiumcitrate, ph 7) containing about 150µg of DNA.

To the sample stabilized at the given temperature (after 5 min. of incubation) 150µl of 10% glyoxal was added and after a further 15 minutes of incubation an additional 150µl of 10% (w/v) glyoxal was added. This mixture was slowly cooled and stored at -20°C until samples could be prepared for electron microscopy using the droplet method.

To the sample containing 0.5-1µg DNA/ml **BAC** (benzyl dimethyl-alkyl-ammonium chloride) was added to a final concentration of 0.0002%. Twenty microliters of this mixture were transformed to fix the DNA in the presence of formaldehyde. The sample was moved onto EM grids and dried twice in 20% (v/v) ethanol.

Samples were shadowed with gold 40% palladium 60% on a rotating stage at a shadowing angle of 15°. The grids were viewed in a **Zeiss 109** electron microscope.

IX.5.2. EM Data Analysis

Photographs of DNA were traced by hand using an XY digitizer connected to a computer (**Tektronix 4051**). The distances between any two points on the molecule and information about the structure of DNA (single or double stranded) was sent to the computer by keys on cursor. This information was stored as a file and could be manipulated and plotted. The data of any set of molecules could be expressed as a histogram.

IX.6. UV Optical Melting Curves

DNA samples from a stock solution were diluted with the following solution : 50mM NaCl, 1mM Tris, 0.5mM EDTA to an absorbance of approximately 0.3 (260nm) in a quartz cuvette (d=5mm). Thermal denaturation was monitored at 260 nm with a programmed temperature increase of 1°C per minute.

Melting curves were continuously recorded using either **Hewlett Packard 8450A Diode Array Spectrophotometer** equipped with a **HP 89100A Temperature Controller** and a **HP 899101A Temperature Station** or

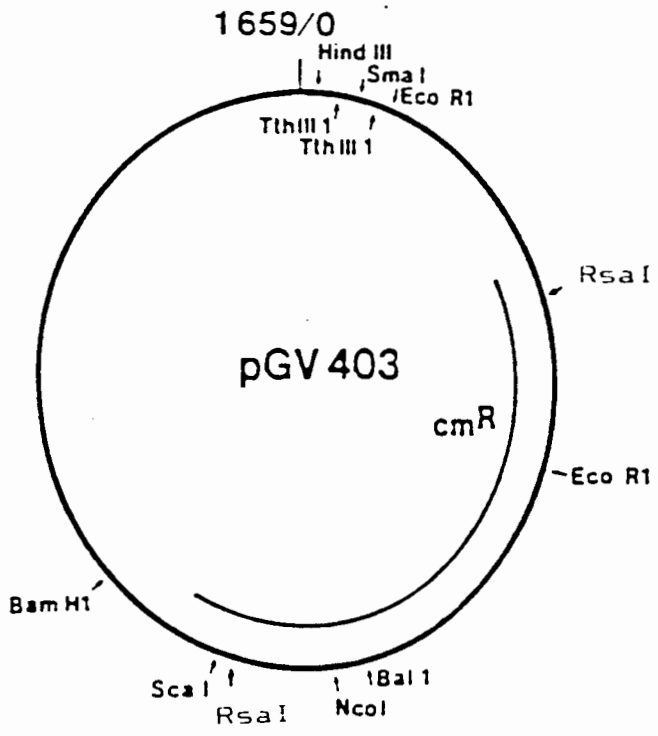
using a **Unicam SP1800 Ultraviolet Spectrophotometer** connected to a water bath. Data were stored on a 3.5 inch disc and plotted using the **HP7470** plotter.

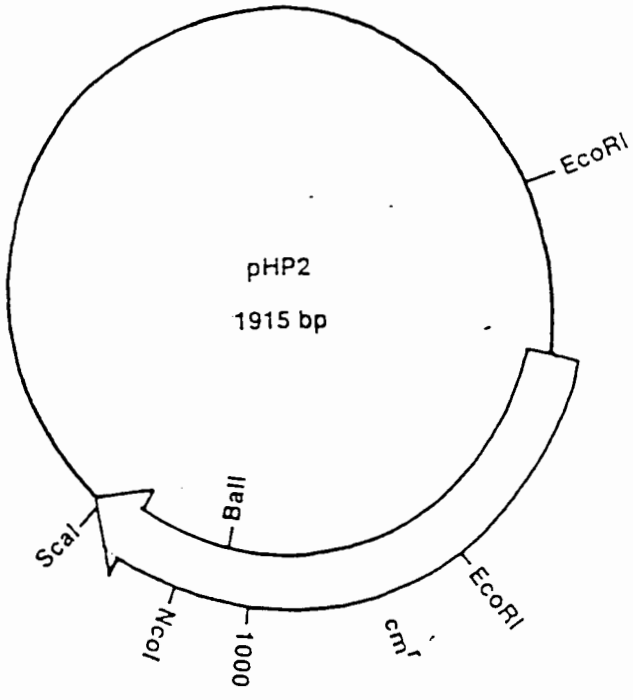
IX.7. Differential scanning calorimetric measurement

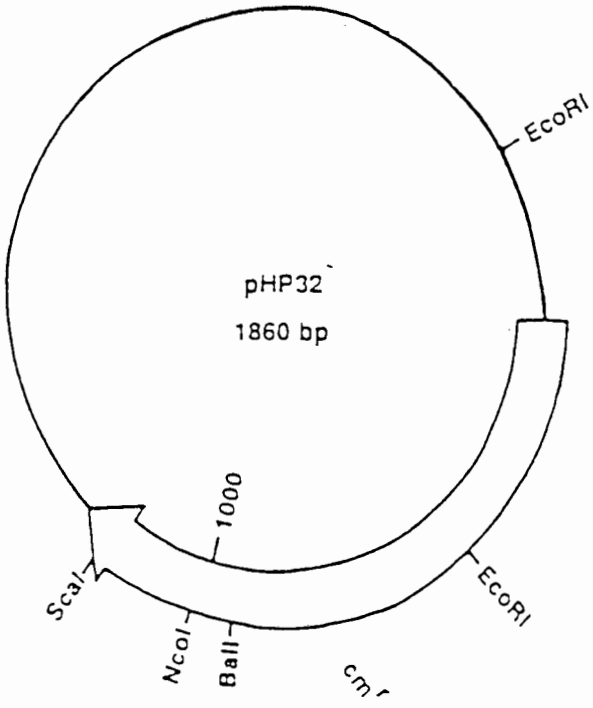
All calorimetric measurements were done using a **DASM-4** Differential Adiabatic Scanning Microcalorimeter equipped with an XY plotter.
Experiments were done over a temperature interval of 70°C (30°C-100°C) with a heating rate of 1°C per minute.
Samples were dissolved in the following buffer (50mM NaCl, 1mM Tris, 0.5mM EDTA) The sample was degassed in order to avoid bubble formation at elevated temperature and was loaded into the calorimetric chambers using a 1ml Hamilton syringe. Each sample was compared with buffer as a reference. DNA concentration was 374μg/ml

Appendix A

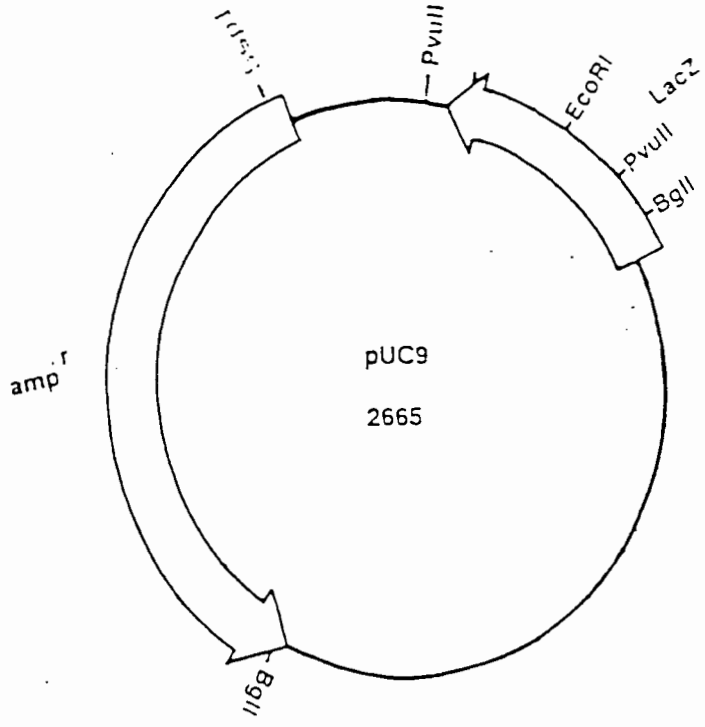
Circular maps of plasmids **pGV403**, **pHP2** and **pHP32**

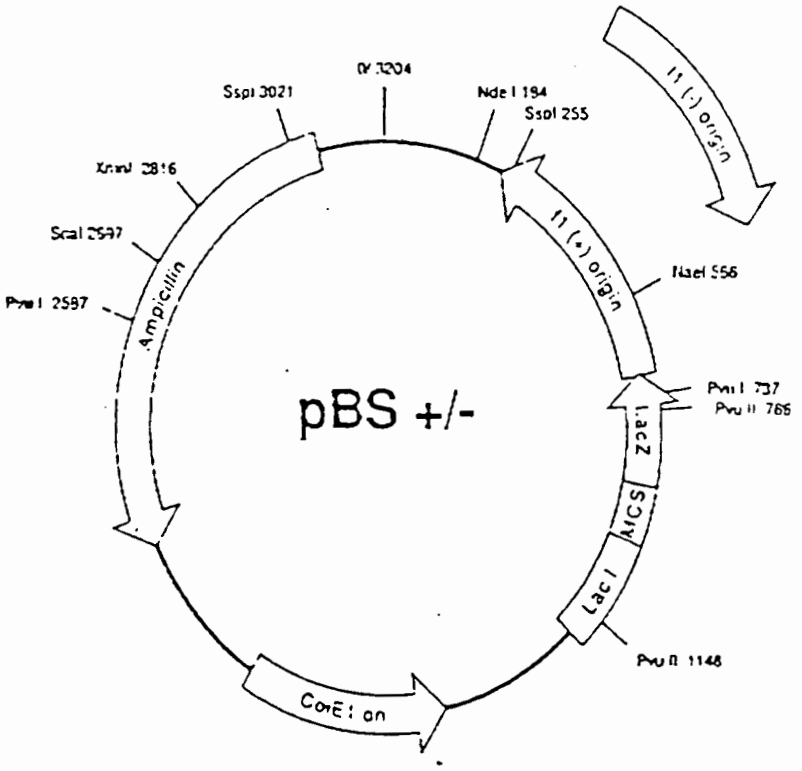


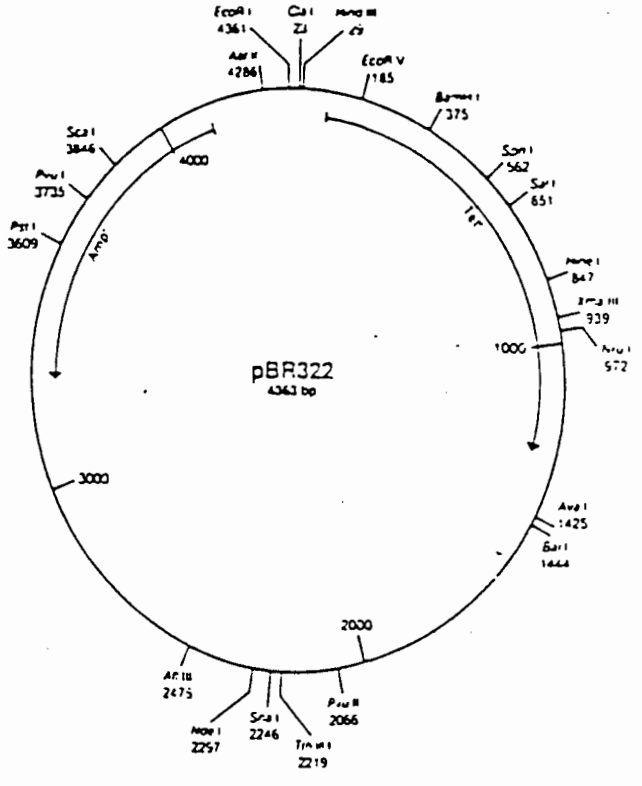




Circular maps of plasmids **pUC9**, **pBS** and **pBR322**




pBS +/-
Sspl 3021
Nde I 194
Sspl 255
f1 (-) origin
f1 (+) origin
Ampicillin
Scal 2597
Pvu I 2587
Pvu II 1148
CoreE1 on
LacI
AlCS
lacZ


pBR322
4363 bp
Amp
Tet
1000
2000
3000
4000

Appendix B

Comparison of sequences: **pGV403** versus **pHP2** and **pGV403** versus **pHP32**

GAP of: Pgp2.Tes check: 3549 from: 1 to: 1915

REFORMAT of: Pgv403.Tes check: 1633 from: 1 to: 1659 August 21, 1991 13:15
(No documentation)

to: Php2.Tes check: 9974 from: 1 to: 1915

REFORMAT of: Php2.Tes check: 9974 from: 1 to: 1915 August 21, 1991 13:16
(No documentation)

Symbol comparison table:
Gencoredisk:[Gcgcore.Data.Rundata]Nwsgapdna.Cmp
CompCheck: 6876

Gap Weight: 5.000 Average Match: 1.000
Length Weight: 0.300 Average Mismatch: 0.000

Quality: 1659.0 Length: 1915
Ratio: 0.866 Gaps: 0
Percent Similarity: 86.632 Percent Identity: 86.632

Pgp2.Tes x Php2.Tes March 17, 1992 10:10 ..

```
                .         .         .         .         .
1 ACGCCAGCAACGCGGCCCGAAGCTTATCGGTGACCCTGACTAAGTCGAGC   50
  ||||||||||||||||||||||||||||||||||||||||||||||||||
1 ACGCCAGCAACGCGGCCCGAAGCTTATCGGTGACCCTGACTAAGTCGAGC   50
                .         .         .         .         .
51 CC************************************************* 100
   ||
51 CCAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCCCAGCTTATAATC 100
                .         .         .         .         .
101 ************************************************** 150

101 ATCCTTATACACGCGCAGTCGATGAGATGAAAAGTTCATTAACGCTACAT 150
                .         .         .         .         .
151 ************************************************** 200

151 TTACAGTGTTTTGGGCAATTCTCCCTCCCCCCCCCCCTCTCTCTCTCTCT 200
                .         .         .         .         .
```

```
201 ************************************************** 250

201 CTCTCTCTCTCTCTCTCCCTTCCTCTAAATATGTTGAATTGAGTGTGT 250

251 ************************************************** 300

251 TTCATGTCAGGAGTGGTATAACTCTCTTTTCCTCTACGTGGGGTACCGAG 300

301 ********GGGCTCGACTCAGTCAGGGCGATGATAAGCTGTCAAACATGA 350
            ||||||||||||||||||||||||||||||||||||||||||
301 CTCGAATTGGGCTCGACTCAGTCAGGGCGATGATAAGCTGTCAAACATGA 350

351 GAATTCTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTA 400
    ||||||||||||||||||||||||||||||||||||||||||||||||
351 GAATTCTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTA 400

401 ATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGA 450
    ||||||||||||||||||||||||||||||||||||||||||||||||
401 ATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGA 450

451 AATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATAT 500
    ||||||||||||||||||||||||||||||||||||||||||||||||
451 AATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATAT 500

501 GTATCCGCTCATGAGACAATAACCCTGATCGAGATTTTCAGGAGCTAAGG 550
    ||||||||||||||||||||||||||||||||||||||||||||||||
501 GTATCCGCTCATGAGACAATAACCCTGATCGAGATTTTCAGGAGCTAAGG 550

501 AAGCTAAAATGGAGAAAAAAATCACTGGATATACCACCGTTGATATATCC 600
    ||||||||||||||||||||||||||||||||||||||||||||||||
551 AAGCTAAAATGGAGAAAAAAATCACTGGATATACCACCGTTGATATATCC 600

601 CAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATG 650
    ||||||||||||||||||||||||||||||||||||||||||||||||
601 CAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATG 650

651 TACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCG 700
    ||||||||||||||||||||||||||||||||||||||||||||||||
651 TACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCG 700

701 TAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCC 750
    ||||||||||||||||||||||||||||||||||||||||||||||||
701 TAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCC 750

751 CGCCTGATGAATGCTCATCCGGAATTCCGTATGGCAATGAAAGACGGTGA 800
    ||||||||||||||||||||||||||||||||||||||||||||||||
751 CGCCTGATGAATGCTCATCCGGAATTCCGTATGGCAATGAAAGACGGTGA 800

801 GCTGGTGATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGC 850
    ||||||||||||||||||||||||||||||||||||||||||||||||
801 GCTGGTGATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGC 850
```

```
851 AAACTGAAACGTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGG 900
    ||||||||||||||||||||||||||||||||||||||||||||||||||
851 AAACTGAAACGTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGG 900

901 CAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCT 950
    ||||||||||||||||||||||||||||||||||||||||||||||||||
901 CAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCT 950

951 GGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCA 1000
    ||||||||||||||||||||||||||||||||||||||||||||||||||
951 GGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCA 1000

1001 ATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGAC 1050
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1001 ATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGAC 1050

1051 AACTTCTTCGCCCCCGTTTTCACCATGGGCAAATATTATACGCAAGGCGA 1100
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1051 AACTTCTTCGCCCCCGTTTTCACCATGGGCAAATATTATACGCAAGGCGA 1100

1101 CAAGGTGCTGATGCCGCTGGCGATTCAGGTTCATCATGCCGTTTGTGATG 1150
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1101 CAAGGTGCTGATGCCGCTGGCGATTCAGGTTCATCATGCCGTTTGTGATG 1150

1151 GCTTCCATGTCGGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAG 1200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1151 GCTTCCATGTCGGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAG 1200

1201 TGGCAGGGCGGGGCGTAATTTTTTTAAGGCAGTTATTGGTGCCCTTAAAC 1250
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1201 TGGCAGGGCGGGGCGTAATTTTTTTAAGGCAGTTATTGGTGCCCTTAAAC 1250

1251 GCCTGGTGCTACGCCTGAATAAGTGATAATAAGCGGATGAATGGCAGAAA 1300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1251 GCCTGGTGCTACGCCTGAATAAGTGATAATAAGCGGATGAATGGCAGAAA 1300

1301 TTCGGATCCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACAC 1350
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1301 TTCGGATCCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACAC 1350

1351 AACATACGAGCCGGAAGCATAAAGTGATAAAGCCTGGGGTGCCTAATGAG 1400
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1351 AACATACGAGCCGGAAGCATAAAGTGATAAAGCCTGGGGTGCCTAATGAG 1400

1401 TGAGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCA 1450
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1401 TGAGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCA 1450

1451 GAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCAC 1500
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1451 GAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCAC 1500
```

```
1501 CACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCT 1550
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1501 CACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCT 1550

1551 GTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGG 1600
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1551 GTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGG 1600

1601 ACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGG 1650
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1601 ACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGG 1650

1651 GGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAG 1700
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1651 GGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAG 1700

1701 ATACCTACAGCGTGAGCATTGAGAAAGCGCCACGCTTCCCGAAGGGAGAA 1750
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1701 ATACCTACAGCGTGAGCATTGAGAAAGCGCCACGCTTCCCGAAGGGAGAA 1750

1751 AGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACG 1800
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1751 AGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACG 1800

1801 AGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTT 1850
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1801 AGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTT 1850

1851 TCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGC 1900
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1851 TCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGC 1900

1901 GGAGCCTATGGAAAA 1915
     |||||||||||||||
1901 GGAGCCTATGGAAAA 1915
```

GAP of: Pgv403.Tes check: 3469 from: 1 to: 1860

REFORMAT of: Pgv403.Tes check: 1633 from: 1 to: 1659 August 21, 1991 13:15
(No documentation)

to: Php32.Tes check: 2654 from: 1 to: 1860

REFORMAT of: Php32.Tes check: 2654 from: 1 to: 1860 August 21, 1991 13:15
(No documentation)

Symbol comparison table:
Gencoredisk:[Gcgcore.Data.Rundata]Nwsgapdna.Cmp
CompCheck: 6876

Gap Weight: 5.000 Average Match: 1.000
Length Weight: 0.300 Average Mismatch: 0.000

Quality: 1659.0 Length: 1860
Ratio: 0.892 Gaps: 0
Percent Similarity: 89.194 Percent Identity: 89.194

Pgv403.Tes x Php32.Tes March 16, 1992 14:10 ..

```
                .         .         .         .         .
1 ACGCCAGCAACGCGGCCCGAAGCTTATCGGTGACCCTGACTAAGTCGAGC 50
  ||||||||||||||||||||||||||||||||||||||||||||||||||
1 ACGCCAGCAACGCGGCCCGAAGCTTATCGGTGACCCTGACTAAGTCGAGC 50
                .         .         .         .         .
51 CC********************************************** 100
   ||
51 CCTCGAATATTAGTAGGAATATGTGCGCGTCAGCTAGTCTACTTTTCAAG 100
                .         .         .         .         .
101 ********************************************** 150

101 TAATTGCGATGTAAATGTCACAAAACCCGTTAAGAGGGAGGGGGGGGGGG 150
                .         .         .         .         .
151 ********************************************** 200

151 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGGGAAGGAGATTTATACAA 200
                .         .         .         .         .
201 ********************************************** 250

201 CTTAACTCACACAAAGTACAGTCCTCACCATATTGAGAGAAAAGGAGATG 250
                .         .         .         .         .
251 ***GGGCTCGACTCAGTCAGGGCGATGATAAGCTGTCAAACATGAGAATT 300
       ||||||||||||||||||||||||||||||||||||||||||||||||
251 CACGGGCTCGACTCAGTCAGGGCGATGATAAGCTGTCAAACATGAGAATT 300
                .         .         .         .         .
```

```
301 CTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTC 350
    ||||||||||||||||||||||||||||||||||||||||||||||||||
301 CTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTC 350

351 ATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGT 400
    ||||||||||||||||||||||||||||||||||||||||||||||||||
351 ATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGT 400

401 GCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATC 450
    ||||||||||||||||||||||||||||||||||||||||||||||||||
401 GCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATC 450

451 CGCTCATGAGACAATAACCCTGATCGAGATTTTCAGGAGCTAAGGAAGCT 500
    ||||||||||||||||||||||||||||||||||||||||||||||||||
451 CGCTCATGAGACAATAACCCTGATCGAGATTTTCAGGAGCTAAGGAAGCT 500

501 AAAATGGAGAAAAAAATCACTGGATATACCACCGTTGATATATCCCAATG 550
    ||||||||||||||||||||||||||||||||||||||||||||||||||
501 AAAATGGAGAAAAAAATCACTGGATATACCACCGTTGATATATCCCAATG 550

551 GCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCT 600
    ||||||||||||||||||||||||||||||||||||||||||||||||||
551 GCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCT 600

601 ATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAG 650
    ||||||||||||||||||||||||||||||||||||||||||||||||||
601 ATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAG 650

651 AAAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCT 700
    ||||||||||||||||||||||||||||||||||||||||||||||||||
651 AAAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCT 700

701 GATGAATGCTCATCCGGAATTCCGTATGGCAATGAAAGACGGTGAGCTGG 750
    ||||||||||||||||||||||||||||||||||||||||||||||||||
701 GATGAATGCTCATCCGGAATTCCGTATGGCAATGAAAGACGGTGAGCTGG 750

751 TGATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAACT 800
    ||||||||||||||||||||||||||||||||||||||||||||||||||
751 TGATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAACT 800

801 GAAACGTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTT 850
    ||||||||||||||||||||||||||||||||||||||||||||||||||
801 GAAACGTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTT 850

851 TCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCT 900
    ||||||||||||||||||||||||||||||||||||||||||||||||||
851 TCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCT 900

901 ATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCC 950
    ||||||||||||||||||||||||||||||||||||||||||||||||||
901 ATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCC 950
```

```
951 TGGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTT 1000
    ||||||||||||||||||||||||||||||||||||||||||||||||||
951 TGGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTT 1000

1001 CTTCGCCCCCGTTTTCACCATGGGCAAATATTATACGCAAGGCGACAAGG 1050
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1001 CTTCGCCCCCGTTTTCACCATGGGCAAATATTATACGCAAGGCGACAAGG 1050

1051 TGCTGATGCCGCTGGCGATTCAGGTTCATCATGCCGTTTGTGATGGCTTC 1100
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1051 TGCTGATGCCGCTGGCGATTCAGGTTCATCATGCCGTTTGTGATGGCTTC 1100

1101 CATGTCGGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCA 1150
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1101 CATGTCGGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCA 1150

1151 GGGCGGGGCGTAATTTTTTTAAGGCAGTTATTGGTGCCCTTAAACGCCTG 1200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1151 GGGCGGGGCGTAATTTTTTTAAGGCAGTTATTGGTGCCCTTAAACGCCTG 1200

1201 GTGCTACGCCTGAATAAGTGATAATAAGCGGATGAATGGCAGAAATTCGG 1250
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1201 GTGCTACGCCTGAATAAGTGATAATAAGCGGATGAATGGCAGAAATTCGG 1250

1251 ATCCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACAT 1300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1251 ATCCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACAT 1300

1301 ACGAGCCGGAAGCATAAAGTGATAAAGCCTGGGGTGCCTAATGAGTGAGG 1350
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1301 ACGAGCCGGAAGCATAAAGTGATAAAGCCTGGGGTGCCTAATGAGTGAGG 1350

1351 ATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCG 1400
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1351 ATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCG 1400

1401 CAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTT 1450
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1401 CAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTT 1450

1451 CAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTAC 1500
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1451 CAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTAC 1500

1501 CAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCA 1550
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1501 CAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCA 1550

1551 AGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTC 1600
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1551 AGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTC 1600
```

```
1601 GTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACC 1650
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1601 GTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACC 1650

1651 TACAGCGTGAGCATTGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCG 1700
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1651 TACAGCGTGAGCATTGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCG 1700

1701 GACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGA 1750
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1701 GACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGA 1750

1751 GCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCC 1800
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1751 GCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCC 1800

1801 ACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGC 1850
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1801 ACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGC 1850

1851 CTATGGAAAA 1860
     ||||||||||
1851 CTATGGAAAA 1860
```

Appendix C

Procedure for calculation of the Local Compositional Complexity

---

```
{************************************************************}
{            procedure reading chosen sequence               }
{global variable bp:number of basepairs in chosen seq        }
{                          a: array of characters,    A,G,C,T}
{************************************************************}

procedure sequence_read{ ( var bp:integer a: tabch ) } ;

var
inputfile:text;

    begin
        textcolor(Blue); highvideo;
        write('give me a file name: ');
        textcolor(yellow);
        repeat
              readln(filename);
        {until filename<>'';}
        assign(inputfile,filename);
        {$I-} reset(inputfile); {SI+}
        bp:=0;
        until (filename<>'') and (IoResult=0 );
        if IoResult = 0 then
        begin
             while not seekeof(inputfile) do
             begin
                  read(inputfile,a[bp]);
                  Inc(bp);
             end;
        end else
        clrscr;
        close(inputfile);
   end; { end of sequence_read }
```

```
{*************************************************************}
{    procedure local compositional complexity                 }
{*************************************************************}

procedure local_complexity;

var
gdriver, gmode              :integer;
xmax, ymax                  :integer;
item, win, wi, j, sh        :integer;
s, bb, s1, mappos, dis      :integer;
n1, n2, n3, n4, increment   :integer;
y11, y12                    :longint;
a1, b1, c1, d1, hg          :real;
h                           :array[1..5000] of real;
st                          :string[4];


      function log(base,number:real):real;

      begin
           if base=0
              then writeln('error: wrong base');
              if number=0
                 then writeln('error: wrong number');
                 if number=1 then log:=0
                 else log:=ln(number)/ln(base);
      end;



   begin
        writeln;
        textcolor(red); write('give me a window size:');
         readln(win);
        detectgraph(gdriver,gmode);
        initgraph(gdriver,gmode,'c:\tp\bgi');
        if graphresult = grOk then
        begin

             wi:=win; j:=1; sh:=1;
             xmax:=getmaxX; ymax:=getmaxY;
             setcolor(red); outtextxy( (xmax-80), 1,
                                       filename);

             n1:=0; n2:=0; n3:=0; n4:=0;

             repeat
                   for i:=sh to win do
                   begin
                        case a[i] of
                             'A','a':inc(n1);
```

```
                              'C','c':inc(n2);
                              'G','g':inc(n3);
                              'T','t':inc(n4);
                         end;
                    end;
                    a1:=n1*log(0.2,n1/wi);
                    b1:=n2*log(0.2,n2/wi);
                    c1:=n3*log(0.2,n3/wi);
                    d1:=n4*log(0.2,n4/wi);
                    h[j]:=(1/wi)*(a1+b1+c1+d1);
                    Inc(sh); Inc(win); Inc(j);
                    n1:=0; n2:=0; n3:=0; n4:=0;
               until win >= bp;

               s:=0; bb:=1;
               s1:=round( j/(xmax-40)+1); hg:=0;

               repeat
                    for i:=1+s to s1+s do
                    begin
                         hg:=hg+h[i];
                    end;
                    h[bb]:=hg/s1; hg:=0; s:=s+s1; inc(bb);
               until bb >= round(j/s1);
               setcolor(magenta); outtextxy(20,ymax-100,'0');
{outtextxy(20,30,'1');}
               setcolor(lightblue); rectangle(30, 30, bb+31,
ymax-100);
               dis:=ymax-130; dis:=round(dis/10);
               setcolor(green);
               for i:=1 to 10 do
               begin
                    str(i:0,st);
                    line(30, (ymax-100)-(i*dis), 34, (ymax-
100)-(i*dis) );
                    settextstyle(2,horizdir,4); outtextxy(1,
(ymax-100)-(i*dis), st);
                    delay(30);
               end;
               settextstyle(0,horizdir,0);
               setcolor(red);  outtextxy(1,ymax-40,'press
Shift-PrtScr if you want to print this screen ');
               setcolor(blue); outtextxy(1,ymax-10,'press Esc
if you want to leave this program ');
               s:=1;
               for i:=2 to bb do
               begin
                    delay(50);
                    setcolor(yellow);
                    yl1:=(ymax-10)-round(sqr(h[i-1])*ymax);
                    yl2:=(ymax-10)-round(sqr(h[i])*ymax);
                    line((i+30),yl1,(i+31),yl2);
                    mappos:=100*s;
```

```
                    if (round(i*s1) > mappos-8) and
(round(i*s1) < mappos+8) then
                    begin
                         if bp>3000 then increment:=4
                         else increment:=2;
                         setcolor(green);
                         line( (i+30),ymax-100,(i+30),ymax-104
);
                         str(mappos,st);
settextstyle(2,horizdir,4); settextjustify(1,1); outtextxy(
(i+30), ymax-90, st);
                         s:=s+increment;;
                    end;


               end;
               settextjustify(centertext, centertext);
settextstyle(2,vertdir, 4); setcolor(green);
               outtextxy(15,(ymax div 2)+50,'local
compositional complexity');
               repeat until readkey=Chr(27);
               closegraph;
          end;
     end; { end of local_complexity }
```

Appendix D

The main aim of this appendix is to show that data compiled for nearest neighbour interaction extracted from the calorimetric measurements are more reliable than the data set calculated from spectrophotometric experiments.

It is possible to distinguish ten nearest neighbor interactions represented by the following combination of basepairs: TA/TA, AA/TT, CA/TG, AT/AT, AG/CT, CG/CG, GA/TC, GG/CC, AC/GT, GC/GC. As mentioned before there are four different datasets in the literature for nearest neighbor interactions, namely Gotoh and Tagashira [41], Breslauer *et. al.* [50], S.G.Delcourt *et. al.*[58] and a new table presented in this thesis obtained from H.Klump {Landolt-Bronstein, Numerical Data and Functional Relationships in Science and Technology, New Series, Group VII, Volume 1, Nucleic Acids, Subvolume c, Spectroscopic and Kinetic Data , Physical Data I p 241]

The relevant data i.e. ΔH and ΔS are presented in tables **14** [41], **15** [51], **16** [61] and **17** respectively.

Table 14

Nearest neighbour thermodynamic parameters according to Gotoh and Tagashira[41]

	ΔH kcal/mol deg	ΔS cal/mol deg
TA/TA	7.43	24.00
AA/TT	7.86	24.00
TG/CA	7.86	24.00
AT/AT	7.92	24.00
AG/CT	7.95	24.00
CG/CG	8.29	24.00
GG/CC	8.61	24.00
GA/TC	8.63	24.00
AC/GT	8.90	24.00
GC/GC	9.82	24.00

Table 15

Nearest neighbour thermodynamic parameters according to Breslauer *et. al.*[50]

	ΔH kcal/mol deg	ΔS cal/mol deg
AA/TT	9.10	24.00
AT/AT	8.60	23.90
TA/TA	6.00	16.90
CA/TG	5.80	12.90
GT/AC	6.50	17.30
AG/CT	7.80	20.80
GA/TC	5.60	13.50
CG/CG	11.90	27.80
GC/GC	11.10	26.70
GG/CC	11.00	26.60

Table 16

Nearest neighbour thermodynamic parameters according to S.G. Delcourt *et. al.*[58]

	ΔH kcal/mol deg	ΔS cal/mol deg
TA/TA	8.187	24.85
AA/TT	8.375	24.85
CA/TG	8.495	24.85
AT/AT	8.548	24.85
AG/CT	8.689	24.85
CG/CG	8.993	24.85
GA/TC	9.011	24.85
GG/CC	9.259	24.85
AC/GT	9.390	24.85
GC/GC	10.141	24.85

Table 17

Nearest neighbour thermodynamic parameters according to H.H.Klump

	ΔH kcal/mol deg	ΔS cal/mol deg
AT/AT	7.46	22.41
TA/TA	7.46	22.41
AA/TT	8.11	23.35
CA/TG	8.47	23.95
AC/GT	8.47	23.95
GA/TC	8.64	23.83
AG/CT	8.64	23.83
GG/CC	9.12	23.49
CG/CG	9.48	25.22
GC/GC	9.48	25.22

Differences can be observed between published data. It is generally assumed that ΔS is constant for all pairs i.e. 24.00 cal/mol*deg for Gotoh and Tagashira[41] and 25.85 cal/mol*deg for Delcourt and Blake[58], whereas data presented by Breslauer et al and data presented in this thesis express ΔS as a function of base combination. Accordingly ΔH for Tables **14** and **16** was calculated using the formula ($\Delta H = T_{MN} * \Delta S$), where ΔS was assumed constant.

Previously presented results based on spectrophotometric and calorimetric experiments suggest that calorimetry could be the best and the most accurate method to establish thermodynamic data for nearest neighbor interactions. The new data set presented in this thesis should be more accurate that those published previously.

REFERENCES

1 E. Schrodinger, *"What is life?"* Cambridge Univ.Press, London and New York **(1945)**

2. C. E. Shannon, *Bell System Techn. J.* **27**, 379 **(1948)**

3. D. Watson and F. Crick, *Nature (London)* **171**, 737 **(1953)**

4. J. Marmur and P. Doty, *J. Mol. Biol.* 5, 109 **(1953)**

5. G. Felsenfeld and G. Sandean, *J. Mol. Biol.* **13**, 407 **(1962)**

6. J. R. Fresco, L. C. Klotz and E. G. Richards, *Cold Spring Harbor Symp. Quant. Biol.* **28**, 83 **(1963)**

7. C. Schildraut and F. Lifson, *Biopolymers* **3**, 195 **(1963)**

8. G. Felsenfeld and S. Z. Hirschman, *J. Mol. Biol.* **28**, 469 **(1967)**

9. D. B. Clewell and D. R. Helinsky, *Proc. Natl. Acad. Sci. USA* **62**, 1159 **(1969)**

10. H. Klump and T. A. Ackermann, *Biopolymers* **10**, 513 **(1971)**

11. W. B. Melchior and P. H. von Hippel, *Proc. Natl. Acad. Sci. U.S.* **70**, 298 **(1973)**

12. F. Michel, J. Lazowska, G. Faye, H. Tukuhara and P. P. Slonimski, *J. Mol. Biol.* **85**, 411 **(1974)**

13. F. Michel, *J. Mol. Biol.* **89**, 305 **(1974)**

14. C. P. Woodbury and M. T. Jr. Record, *Biopolymers* 14, 2417 **(1975)**

15. S. Yabuki, O. Gotoh and A. Wada, *Biochim. Biophys. Acta* **395**, 258 (1975)

16. P. L. Privalov, V. V. Plotnikov and V. V. Filimonov, *J. Chem. Thermodynamics* **7**, 41 (**1975**)

17. P. Botchan, *J. Mol. Biol.* **105**, 161 (**1976**)

18. A. Wada, H. Tachibana, O. Gotoh and M. Takanami, *Nature* **263**, 439 (**1976**)

19. O. Gotoh, Y. Husimi, S. Yabuki and A. Wada, *Biopolymers* **15**, 655 (**1976**)

20. D. L. Vizard and A. T. Ansevin, *Biochemistry* **15**(4), 741 (**1976**)

21. V. M. Pavlov, Yu. L. Lyubchenko, A. S. Borovik and Yu. S. Lazurkin, *Nucl. Acids. Res.* **4**, 4053 (**1977**)

22. H. Klump and W. Burkart, *Biochim. Biophys. Acta* **475**, 601 (**1977**)

23. H. Klump, *Biochim. Biophys. Acta* **475**, 605 (**1977**)

24. S. Dasgupta, D. P. Allison, C. E. Srigler and S. Mitra, *J. Biol. Chem.* **252**, 5916 (**1977**)

25. R. M. Wartell, *Natl. Acad. Res.* **4**, 2779 (**1977**)

26. B. B. Jones, H. W. Chen, S. Rothstain, R. D. Wells and W. S. Resnikoff, *Proc. Natl. Acad. Sci.* **74**, 4914 (**1977**)

27. A. Wada, H. Tachibana, S. Ueno, Y. Husimi and Y.Machida, *Nature* **269**, 352 (**1977**)

28. C. Takiyama, O. Gotoh and A. Wada, *Biopolymers* **16**, 427 (**1977**)

29. R. D. Blake and S. G. Lefoley, *Biochim. Biophys. Acta* **518**, 2623 (**1978**)

30. J. G. Sutcliffe, *Cold Spring Harbor Symp. Quant. Biol.* **43**, 77 (**1978**)

31. J. G. Sutcliffe, *Proc. Natl. Acad. Sci.* **75**, 3737 (**1978**)

32. B. Y. Tong and S. J. Battersby, *Biopolymers* **17**, 2933 (**1978**)

33. H. Tachibana, A. Wada, O. Gotoh and M. Takanami, *Biochem. Biophys. Acta* **517**, 319 (**1978**)

34. A. Wada, S. Ueno, H. Tachibana and Y. Husimi, *J. Biochem.* **85**, 827 (**1978**)

35. E. Freire and R. L. Biltonen, *Biopolymers* **17**, 481 (**1978**)

36. S. C. Hardies, W. Hillen, T. C. Goodman and R. D. Wells, *J.Biol.Chem.* **254**, 10128 (**1979**)

37. O. Gotoh, A. Wada and S. Yabuki, *Biopolymers* **18**, 805 (**1979**)

38. H. J. Vollenweider, M. Fiandt and W. Szybalski, *Science* **205**, 508 (**1979**)

39. A.S.Borovik, A.S.Kalambet, Yu.L.Lyubchenko, V.Tishkov and Eu.I.Gdoranov, *Nucl. Acids. Res.* **8**, 4165 (**1980**)

40. A. Wada, S. Yabuki and Y. Husimi, *CRC Crit. Rev. Biochem* **9**, 87 (**1980**)

41. O. Gotoh and Y. Tagashira, *Biopolymers* **20**, 1033 (**1981**)

42. O. Gotoh, *Adv. Biophys.* **16**, 1 (**1983**)

43. J. Vieire and J. Mesing, *Gene* **19**, 259 (**1982**)

44. T. Maniatis, E. F. Fritsch and J. Sambrook, Molecular cloning: *A laboratory manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.* (**1982**)

45. K. W. C.Peden, *Gene* **22**, 277 (**1983**)

46. Y. Maeda, Y. Kawai,T. Fujita and E. Othsubo, *J. Gen. App. Microbiol.* 30, 289 (**1984**)

47. Y. Maeda, Y. Kawai, T. Fujita and E. Othsubo, *Termochim. Acta* **88**, 235 (**1985**)

48. Yu. A. Kalambet, A. S. Borovik, V. I. Lyamichev and Yu. L. Lyubchenko, *Biopolymers* **24**, 235 (**1985**)

49. R. D. Blake and T. G. Hydorn, *J. Biochem. Biophys. Meth.* **11**, 307 (**1985**)

50. K. J. Breslauer, R. Frank, H. Bloker and L. A. Marky, *Proc. Natl. Acad. Sci. USA* **83**, 3746 (**1986**)

51. A. Wada and A. Suyama, *Progr. Biophys. Met. Biol.* **47**, 113 (**1986**)

52. P. L. Privalov and S. A. Potekhin, *Meth. Enzymology* **131**, 4 (**1986**)

53. Y. Maeda and E. Othsubo, *J. Mol. Biol* **194**, 691 (**1987**)

54. R. D. Blake and S. G. Delkourt, *Biopolymers* **26**, 2009 (**1987**)

55. L. Marky and K. J. Breslauer, *Biopolymers* **26**, 1601 (**1987**)

56. G. Bernardi, D. Mouchiroud, C. Gautier and G. Bernardi, *J. Mol. Evol.* **28**, 7 (**1988**)

57. A. Konopka and J. Owens, *Gene. Anal. Tech. App.* **7**, 35 (**1990**)

58. S. G.Delcourt and R. D.Blake, *J. Biol. Chem.* **266**, 151 (**1991**)

59. Batch analysis sheet, Amersham Internatioal plc

60. V. Brendel, *Ph.D. Thesis, Weizmann Institute of Science, Rehovot, Israel "Towards an understanding of DNA as a Language"*

61. S. A. Potekhin and P. L. Privalov, *J. Mol. Biol.* **159**, 519 (**1982**)

62. R. L. Dorit, L. Schoenbach and W. Gilbert, *Science* **250**, 1377 (**1990**)